UNIVERSITY OF CALIFORNIA, MERCED

Soil Structure and Land Surface Controls on Soil Hydraulic Properties and Processes: Applications of Machine Learning, Unmanned Aircraft Systems, and Observations from Long-Term Conservation Agriculture Management

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Environmental Systems

by

Samuel Negusse Araya

Committee in charge:

Professor Teamrat A. Ghezzehei, Chair
Professor YangQuan Chen
Professor Martha Conklin
Professor Jason Sexton
Professor Joshua H. Viers

2019

The dissertation of Samuel Negusse Araya is approved, and is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, Merced

2019

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to briefly thank some of the persons without whose support, guidance, and generosity I would not have been able to complete my dissertation.

My mentor, Professor Teamrat Afewerki Ghezzehei, for his support and guidance.

My dissertation committee members: Professors YangQuan Chen, Martha Conklin, Jason Sexton, and Joshua H. Viers for their help and insights.

Doctor Yakov Pachepsky (US Department of Agriculture) for generously providing the database which was the basis for the first section of my dissertation.

Doctor Attila Nemes (Norwegian Institute of Bioeconomy Research), and three anonymous reviewers for the Journal of Water Resources Research. Their comments and feedback have greatly improved the first section of my dissertation.

Professor Jeffrey P. Mitchell (UC Davis) for his support, mentorship, and expertise that made it possible to study soils at the UC Agriculture and Natural Research site at Five Point, CA, and Professor Jan Hopmans (UC Davis) for his valuable input and comments which greatly benefited the second part of my dissertation research.

Anna Fryjoff-Hung and Andreas Anderson, whose expertise and collaboration enabled me to conduct the third part of my research.

The MERCED high performance computing cluster staff, Sarvani Chadalapaka and Matthias Bussonnier, for their generous help in accessing and using the MERCED Cluster.

Professor Asmeret Asefaw Berhe for her mentorship and support.

On the social and personal side,

first and foremost, I would like to thank my wife Yirgalem Debessai Araya, for her support, patience, and always being there for me.

My parents, Negusse and Gebriela, and all my siblings for making me who I am and always helping me to be a better person.

My church family at Creekside, Merced.

Fellow graduate students and colleagues at the UC Merced Soil labs.

To all friends and family, too many to list here but whose friendship and support I will never forget.

I am ever thankful to God for His love and for surrounding me with all those wonderful people.

This work would not have been possible without the support from the School of Natural Sciences and School of Engineering at UC Merced.

Vita

Samuel Negusse Araya

EDUCATION

- Doctor of Philosophy candidate in Environmental Systems
  University of California at Merced, 2014 – present.
  Dissertation title: "Soil Structure and Land Surface Controls on Soil Hydraulic Properties and Processes: Applications of Machine Learning, Unmanned Aircraft Systems, and Observations from Long-Term Conservation Agriculture Management."

- Master of Science in Environmental Systems
  University of California at Merced, 2010 – 2014.
  Thesis title: "Effect of Combustion Temperature on Soil and Soil Organic Matter Properties: A Study of Soils from the Western Elevation Transect in Central Sierra Nevada, California."

- Bachelor of Science in Land Resources and Environment
  University of Asmara 2002 – 2006.

PROFESSIONAL AND ACADEMIC EMPLOYMENT

- Teaching Assistant, University of California at Merced. Jan. 2011 – May. 2019.
  Courses taught: Stable Isotope Ecology (BIO 174/ESS 174) • Fundamentals of Soil Science (ES 170) • Contemporary Biology (BIOL 01) • Earth Resources and Society (ESS 010) • Sustainability Science (ESS 002).

- Graduate Teaching Assistant, Hamelmalo Agricultural College, Eritrea. Aug. 2006 – Aug. 2010.
  Courses taught: Fundamentals of Soil Science (LREN212D) • Fundamentals of Remote Sensing and GIS (LREN241) • Land Use Planning (LREN202).

- Academic Programming Director, Hamelmalo Agricultural College, Eritrea Mar. 2007 – Aug. 2010.

PUBLICATIONS

Peer-reviewed

- Araya, S. N., & Ghezzehei, T. A. (2019). Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and its Sensitivity to Soil Structural Perturbation. Water Resources Research. 55(7). 5715-5737. doi:10.1029/2018WR024357

- Araya S. N., Fogel L. M. & Berhe A. A. (2017). Thermal Alteration of Soil Organic Matter Properties: A Systematic Study to Infer Response of Sierra Nevada Climosequence Soils to Forest Fires. SOIL. 3(1). 31-44. doi: 10.5194/soil-3-31-2017.
- Araya S. N., Meding M. & Berhe A. A. (2016). Thermal Alteration of Soil Physico-Chemical Properties: A Systematic Study to Infer Response of Sierra Nevada Climosequence Soils to Forest Fires. SOIL. 2(3). 351–366. doi: 10.5194/soil-2-351-2016.
- Berhe A. A., Arnold C., Stacy E., Lever R., McCorkle E. & Araya S. N. (2014). Soil Erosion Controls on Biogeochemical Cycling of Carbon and Nitrogen. Nature Education Knowledge. 5(8):2

Manuscripts in preparation

- Araya, S. N., Fryjoff-Hung, A., Anderson, A., Viers, J. H., & Ghezzehei, T. A. Prediction of Soil Moisture at High Spatial Resolution Using Machine Learning and Unmanned Aircraft Systems-Based Remote Sensing.
- Araya, S. N., Mitchell, J. P., Hopmans, J. H. & Ghezzehei, T. A. Long-Term Impact of Conservation Agriculture on Soil Hydraulic Properties and Dynamic Soil Moisture Storage

CONFERENCE PRESENTATIONS

- Araya, S. N., Mitchell J. P., and Ghezzehei T. (2019). How 20-Years of Conservation Agriculture Modified Soil Hydraulic Properties: Lab Analysis and Numerical Simulations. Poster presented at the 2019 International Soils Meeting, San Diego, CA, 6-9 January.
- Araya, S. N., et al. (2018). Estimating Soil Moisture from UAV and Machine Learning. Poster presented at the 2018 AGU Fall Meeting, Washington, DC, 10-14 December.
- Araya, S. N. and Ghezzehei T. (2018). Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and its Sensitivity to Soil Structural Perturbations. Poster presented at the 2018 AGU Fall Meeting, Washington, DC, 10-14 December.
- Araya, S. N., et al. (2018). Estimating Soil Moisture from Unmanned Aerial Vehicle and Machine Learning. Poster presented at the Inaugural Data Science Workshop, Data Science Institute, Lawrence Livermore National Laboratory, Livermore, 7-8 August.
- Araya, S. N. and Ghezzehei T. (2018). Machine Learning to Predict the Effect of Soil Structure on Saturated Hydraulic Conductivity. Oral presentation at the W-

3188: Soil, Water, and Environmental Physics Across Scales Meeting, Las Vegas, NV, 2-4 January.

- Araya, S. N. and Ghezzehei T. (2017). Using Machine Learning to Predict the Effect of Soil Structure on Saturated Hydraulic Conductivity. Poster presented at the 2017 AGU Fall Meeting, New Orleans, LA, 11-15 December.
- Araya, S. N., Berli M. and Ghezzehei T. (2017). Soil moisture recharge estimation: a mass balance approach for quantifying effective precipitation. Poster presented at the 26th Groundwater Resources Association of California Annual Meeting, Sacramento, CA, 3-4 October.
- Araya, S. N. (2016). Exploring the effectiveness of active learning methods in a discussion class of an introductory environmental science course. Poster presented at the Assessment as Research Symposium, UC Merced, 2 March.
- Araya, S. N., Berhe A. (2014). Effect of Combustion Temperature on Soil and Soil Organic Matter Properties: A Study of Soils from the Western Elevation Transect in Central Sierra Nevada, California. Poster presented at the 2014 AGU Conference, San Francisco, CA, 15-19 December.

PROFESSIONAL MEMBERSHIPS:

- American Geophysical Union (AGU)
- American Society of Agronomy (ASA)
- Crop Science Society of America (CSSA)
- European Geosciences Union (EGU)
- Soil Science Society of America (SSSA)

Samuel Negusse Araya
Environmental Systems
University of California, Merced, 2019
Major Professor: Dr. Teamrat A. Ghezzehei

Soil Structure and Land Surface Controls on Soil Hydraulic Properties and Processes: Applications of Machine Learning, Unmanned Aircraft Systems, and Observations from Long-Term Conservation Agriculture Management

General Abstract

Soil moisture exerts a strong influence on ecology and energy balance of the environment. Its high variability in space and time, however, makes it difficult to measure accurately. Predicting the dynamics of soil moisture is also challenging because of the complex relationship between soil hydraulic properties and other physical characteristics. In this study, the links between soil structure and land surface characteristics with soil hydrology were determined using machine learning, unmanned aircraft system, and observations from long-term conservation agriculture management.

The saturated hydraulic conductivity is an essential property of soil that determines the fate of soil moisture. Machine learning models were developed that predict saturated hydraulic conductivity with significantly improved accuracy than previous models. The impact of soil structural changes on hydraulic conductivity was further investigated using these models. Saturated hydraulic conductivity showed an inverse relationship with bulk density. The most critical change occurred around a bulk density of 1.8 g cm^{-3}. The relationship between organic carbon and saturated hydraulic conductivity showed a positive relationship for fine- and mid-texture soils but an inverse relation for coarse-textured soils. To analyze the impact of soil structural change on a more comprehensive set of soil hydraulic properties and water dynamics, the same set of soils that have been under different tillage and cover crop systems for 18-years were studied, and the effects on soil hydraulic properties and processes quantified. Conventional measures of soil structure and hydraulic properties, as well as numerical simulations, were applied to investigate the implications on soils' capacity to capture and store water. Combined reduced tillage and cover cropping practices led to an improvement in soil structure and soils ability to retain more water when analyzed by numerical simulations. Conventional, static, measures of water contents at field capacity and plant available water contents, however, suggested a decrease in soils' capacity to hold water in the reduced tillage and cover crop plots. The findings suggest that conventional static measures of soil hydraulic property may not be adequate to explain soils' ability to capture and retain water. To investigate the relationship between surface characteristics and soil water dynamics at the landscape

scale, machine learning models were trained to interpret surface soil moisture across a grassland landscape from surface characteristics recorded by a multispectral camera from unmanned aircraft system platform. The models predicted surface soil moisture with reasonable accuracy. Analysis of variable importance revealed that reflectance in the red band and topographic position indices to be the two most important predictors outside of weather variables. By using modeling, machine learning, data fusion, and unmanned aircraft systems, this dissertation investigated and described the relationship of soil hydraulic properties and processes with soil structural and land surface characteristics.

Chapter 1 Summary Introduction

The relatively small quantity of water stored in the upper layers of soil plays a crucial role in terrestrial biology, biogeochemistry, and atmospheric water and energy fluxes. More than half of the solar energy absorbed by the land surface is used to evaporate water (Trenberth et al., 2009) and about 60% of terrestrial precipitation is returned to the atmosphere by evapotranspiration (Seneviratne et al., 2010). In many ecosystems soil moisture is the most important resource affecting vegetation structure and organization, thereby determining primary productivity, nutrient mineralization, and the ensuing biogeochemistry (Rodriguez-Iturbe et al., 2001). Biogeochemical heterogeneity in soils are also closely linked to the variation in soil moisture (Wanzek et al., 2018).

Soil moisture is highly variable in space and time, and it is difficult to measure accurately. Predicting the dynamics of soil moisture is also challenging because the relation between soil moisture and the factors that drive its dynamics are very complex. In most environments, soil water storage mainly depends on precipitation and evapotranspiration (Hillel, 1998; Rana & Katerji, 2000). The distribution of water in the soil is dependent on the soils' hydraulic properties, topography, and other surface and subsurface properties (Sørensen et al., 2006; Vereecken et al., 2016). The immediate dynamics of soil moisture is mainly driven by the soil hydraulic properties, which in turn are the result of several soil characteristics.

Two fundamental attributes that control the soil water dynamics are the soil's water retention and conductivity property. Water inside the soil is held at a negative pressure (i.e., at a suction) that increases rapidly as soils dry. The relation between water content and the suction pressure at which water is held inside a soil is described by the water retention curve (Figure 1-1A). Similarly, the relation between the hydraulic conductivity of soils and the suction pressure is described by the hydraulic conductivity function (Figure 1-1B). The hydraulic conductivity of soils decreases very rapidly as soils dry, and the highest conductivity for a soil occurs when the soil is saturated.

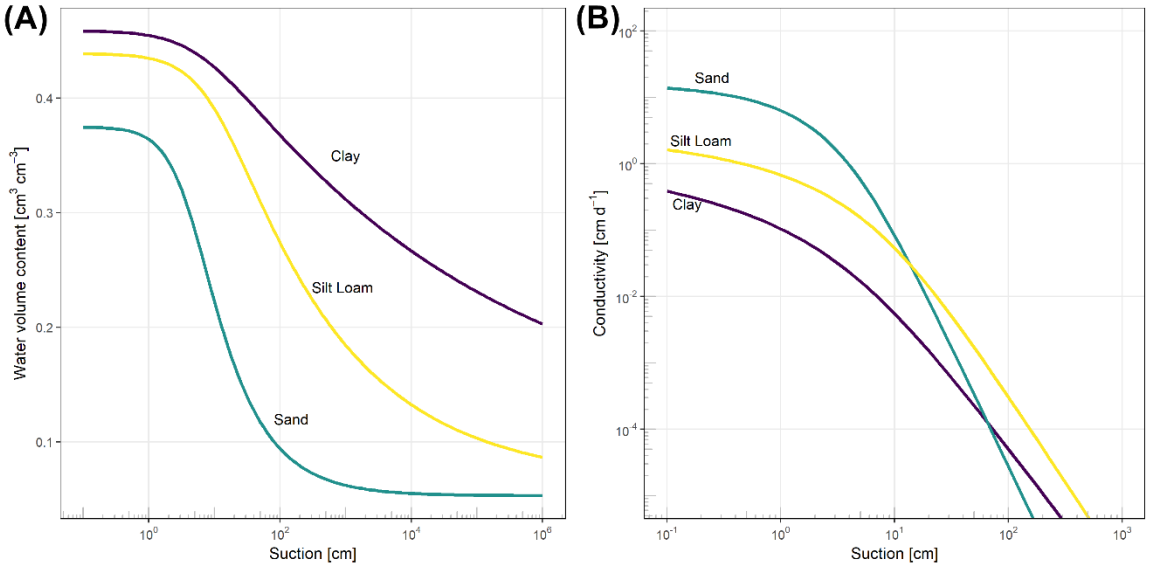

Figure 1-1 Typical water retention (A) and hydraulic conductivity (B) of three soils versus suction (suction in units of hydraulic head).

Soil structure—the arrangement and packing of soil particles—is an important factor that controls soil hydraulic properties and unlike soil texture it is a highly dynamic property that can change in a short amount of time in response to natural and management conditions (Assouline & Or, 2013; de Almeida et al., 2018; Jorda et al., 2015). Quantitatively understanding the relationship between soil physical properties and the soil hydraulic properties is crucial in enabling the prediction of soil moisture dynamics.

Many soil hydraulic properties are often estimated from other, more easily measured, soil properties using a variety of predictive models known as pedotransfer functions (PTFs) (Bouma, 1989). Predictions from PTFs can be used to extrapolate knowledge on soil properties from one location to another (Rahmati et al., 2018) or as input into simulation models to run scenarios on the effect of management and other environmental processes (José Padarian et al., 2018). PTF predictions of soil hydraulic properties are particularly common because of the difficulty in measuring those properties sufficiently in time and space, and the utility of hydraulic properties in many applications including in land surface modeling (Chaney et al., 2019; Van Looy et al., 2017). Including soil structural variables as predictors in PTFs allows the capture of the dynamics as a result of changes in soil structure and informs on the relationship between soil structure and hydraulic properties.

Traditionally, PTFs are developed from point source data and are non-spatial. Most PTFs are not readily adapted for spatial prediction (Hengl et al., 2018). Environmental parameters of topography and vegetation have been shown to improve spatial PTF predictions of soil hydraulic properties (Jana & Mohanty, 2011; Pachepsky et al., 2001; J. Padarian et al., 2019; Sharma et al., 2006). Padarian et al. (2018) suggest that models that depend only on non-soil environmental information are not technically PTFs but spatial prediction functions. Both the traditional PTFs and the so-called spatial prediction functions suffer from the challenge of the relation between the observed inputs and soil

hydraulic properties being very complex. It is challenging to model these relationships accurately using physically-based models or traditional statistical methods. This opens an opportunity to revisit these challenges using data-driven methods which excel in such situations, such as machine learning techniques and high-resolution data (Shen et al., 2018).

This study was motivated by the growing availability of large databases, modern data acquisition techniques such as unmanned aircraft systems, and increasing computing power enabling data intensive methods. These advances make the problem of quantifying the relationship between physical properties and its hydraulic properties a good candidate for methods that employ modeling, the fusion of data from different sources and scales, and machine learning-based investigations.

1.1 General Objectives

The overarching objectives of the study which are covered in three separate chapters are to:

1. describe how soil structure influences saturated hydraulic conductivity,
2. determine changes in soil hydraulic properties and processes that result from management, and
3. determine how surface soil moisture distribution responds to landscape-scale surface variables.

A theme that runs across all the three chapters is modeling, data fusion, and machine learning

1.2 Organization of the Dissertation

The investigations in this dissertation are organized into three main sections and a general conclusion.

In chapter 2, machine learning models were developed to predict the saturated hydraulic conductivity of soils and determine the responses to structural alterations.

In chapter 3, the implication of soil structural changes resulting from long-term conservation agriculture management on soil hydraulic properties and soil water storage was analyzed.

In chapter 4, surface soil moisture distribution across a landscape was described though machine learning assisted interpretation of surface characteristics observed from an unmanned aircraft system platform.

In chapter 5, the main findings are summarized.

References

Assouline, S., & Or, D. (2013). Conceptual and Parametric Representation of Soil Hydraulic Properties: A Review. *Vadose Zone Journal*, *12*(4), 0. https://doi.org/10.2136/vzj2013.07.0121

Bouma, J. (1989). Using Soil Survey Data for Quantitative Land Evaluation. In B. A. Stewart (Ed.), *Advances in Soil Science* (pp. 177–213). New York, NY: Springer US. https://doi.org/10.1007/978-1-4612-3532-3_4

Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., … Yimam, Y. (2019). POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties Over the Contiguous United States. *Water Resources Research*, *55*(4), 2916–2938. https://doi.org/10.1029/2018WR022797

de Almeida, W. S., Panachuki, E., de Oliveira, P. T. S., da Silva Menezes, R., Sobrinho, T. A., & de Carvalho, D. F. (2018). Effect of soil tillage and vegetal cover on soil water infiltration. *Soil and Tillage Research*, *175*(June 2017), 130–138. https://doi.org/10.1016/j.still.2017.07.009

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. https://doi.org/10.7717/peerj.5518

Hillel, D. (1998). *Environmental Soil Physics*. San Diego, CA: Academic Press.

Jana, R. B., & Mohanty, B. P. (2011). Enhancing PTFs with remotely sensed data for multi-scale soil water retention estimation. *Journal of Hydrology*, *399*(3–4), 201–211. https://doi.org/10.1016/j.jhydrol.2010.12.043

Jorda, H., Bechtold, M., Jarvis, N., & Koestel, J. (2015). Using boosted regression trees to explore key factors controlling saturated and near-saturated hydraulic conductivity. *European Journal of Soil Science*, *66*(4), 744–756. https://doi.org/10.1111/ejss.12249

Pachepsky, Y. A., Timlin, D. J., & Rawls, W. J. (2001). Soil Water Retention as Related to Topographic Variables. *Soil Science Society of America Journal*, *65*(6), 1787. https://doi.org/10.2136/sssaj2001.1787

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional*, *16*, e00198. https://doi.org/10.1016/j.geodrs.2018.e00198

Padarian, José, Morris, J., Minasny, B., & McBratney, A. B. (2018). Pedotransfer Functions and Soil Inference Systems. In A. McBratney, B. Minasny, & U. Stockmann (Eds.), *Pedometrics. Progress in Soil Science.* (pp. 195–220). Springer, Cham. https://doi.org/10.1007/978-3-319-63439-5_7

Rahmati, M., Weihermüller, L., Vanderborght, J., Pachepsky, Y. A., Mao, L., Sadeghi, S. H., … Vereecken, H. (2018). Development and analysis of the Soil Water Infiltration Global database. *Earth System Science Data*, *10*(3), 1237–1263. https://doi.org/10.5194/essd-10-1237-2018

Rana, G., & Katerji, N. (2000). Measurement and estimation of actual evapotranspiration in the field under Mediterranean climate: a review. *European Journal of Agronomy*, *13*(2–3), 125–153. https://doi.org/10.1016/S1161-0301(00)00070-8

Rodriguez-Iturbe, I., Porporato, A., Laio, F., & Ridolfi, L. (2001). Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress. *Advances in Water Resources*, *24*(7), 695–705. https://doi.org/10.1016/S0309-1708(01)00004-5

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., … Teuling, A. J. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Sharma, S. K., Mohanty, B. P., & Zhu, J. (2006). Including Topography and Vegetation Attributes for Developing Pedotransfer Functions. *Soil Science Society of America Journal*, *70*(5), 1430. https://doi.org/10.2136/sssaj2005.0087

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F. J., … Tsai, W. P. (2018). HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. *Hydrology and Earth System Sciences*, *22*(11), 5639–5656. https://doi.org/10.5194/hess-22-5639-2018

Sørensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: Evaluation of different methods based on field observations. *Hydrology and Earth System Sciences*, *10*(1), 101–112. https://doi.org/10.5194/hess-10-101-2006

Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's Global Energy Budget. *Bulletin of the American Meteorological Society*, *90*(3), 311–323. https://doi.org/10.1175/2008BAMS2634.1

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., … Vereecken, H. (2017). Pedotransfer Functions in Earth System Science: Challenges and Perspectives. *Reviews of Geophysics*, *55*(4), 1199–1256. https://doi.org/10.1002/2017RG000581

Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., … Young, I. M. (2016). Modeling Soil Processes: Review, Key Challenges, and New Perspectives. *Vadose Zone Journal*, *15*(5), 0. https://doi.org/10.2136/vzj2015.09.0131

Wanzek, T., Keiluweit, M., Baham, J., Dragila, M. I., Fendorf, S., Fiedler, S., … Kleber, M. (2018). Quantifying biogeochemical heterogeneity in soil systems. *Geoderma*, *324*(September 2017), 89–97. https://doi.org/10.1016/j.geoderma.2018.03.003

Chapter 2 Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations[1]

Abstract

Saturated hydraulic conductivity (K_s) is a fundamental soil property that regulates the fate of water in soils. Its measurement, however, is cumbersome and instead pedotransfer functions (PTFs) are routinely used to estimate it. Despite much progress over the years, the performance of current generic PTFs estimating K_s remains poor. Using machine learning, high-performance computing, and a large database of over 18,000 soils, we developed new PTFs to predict K_s. We compared the performances of four machine learning algorithms and different predictor sets. We evaluated the relative importance of soil properties in explaining K_s. PTF models based on boosted regression tree algorithm produced the best models with root mean squared log-transformed error in ranges of 0.4 to 0.3 ($\log_{10}(cm/day)$). The 10th percentile particle diameter (d_{10}) was found to be the most important predictor followed by clay content, bulk density (ρ_b), and organic carbon content (C). The sensitivity of K_s to soil structure was investigated using ρ_b and C as proxies for soil structure. An inverse relationship was observed between ρ_b and K_s, with the highest sensitivity at around 1.8 g cm^{-3} for most textural classes. Soil C showed a complex relationship with K_s with an overall positive relation for fine- and mid-textured soils but an inverse relation for coarse-textured soils. This study sought to maximize the extraction of information from a large database to develop generic machine learning-based PTFs for estimating K_s. Models developed here have been made publicly available and can be readily used to predict K_s.

[1] Chapter published in a journal: Araya, S. N. and Ghezzehei, T. A. (2019). Water Resources Research. doi:10.1029/2018WR024357.

List of Acronyms and Symbols

BRT	Boosted regression trees
KNN	K-nearest neighbors
ML	Machine learning
MLE	Mean log-transformed error
ML-PTF	Machine learning-based pedotransfer function
PTF	Pedotransfer function
RF	Random forest
RMSLE	Root mean squared log-transformed error
SVR	Support vector regression

ρ_b	Bulk density
C_X	Complexed organic carbon
C	Organic carbon
K_s	Saturated hydraulic conductivity

2.1 Introduction

Hydraulic conductivity of water-saturated soils (K_s) is one of the most important soil characteristic that determines rate of infiltration, runoff generation, and deep drainage. Its importance is particularly elevated during precipitation, snowmelt, flooding, and/or irrigation events. K_s regulates the amount of plant-available water, overland flow and transport, erosion, groundwater recharge, and extent and duration of water inundation. The magnitude of soil hydraulic conductivity primarily depends on the size, distribution and, connectivity of pores (Alaoui et al., 2011; Bittelli et al., 2015; Nielsen et al., 2018). Thus the first order classification of soil hydraulic conductivity classes is typically dependent upon soil texture, ranging from > 5 m/day for sandy soils to < 0.01 m/day for clay textured soils (Rawls et al., 1982; e.g., Soil Science Division Staff, 2017). In addition to texture, hydraulic conductivity is influenced by the soil structure which itself is the result of several factors. Structure determines the presence and connectivity of large pores—including macropores, cracks, and inter-aggregate pore spaces (Beven & Germann, 1982).

Unlike texture, soil structure is prone to substantial alteration in a relatively short time, which could have consequential effects on hydraulic conductivity and associated hydrologic processes (Assouline & Or, 2013). Common structure altering processes that have strong bearing on hydraulic conductivity include burrowing by roots and soil fauna, aggregation, compaction, and wetting/drying cycles (Brooks et al., 2004; Chivenge et al., 2007; Ghezzehei & Or, 2000; Kuncoro et al., 2014; Or et al., 2000). These processes may occur on seasonal cycles or over several decades as part of soil development. More drastic changes to soil structure such as tillage and cracking can increase hydraulic conductivity by several orders of magnitude, albeit for a short period of time (de Almeida et al., 2018; Jorda et al., 2015).

It is often impractical to measure hydraulic conductivity with adequate spatial density and frequency as soil hydraulic properties vary considerably across landscapes, often within short distances. Lack of adequate information that captures the spatial heterogeneity and temporal dynamics of soil hydraulic conductivity and related processes are often identified as critical shortcomings in land surface models that simulate processes across large regions and long periods. In this regard, pedotransfer functions (PTF)—models that predict soil hydraulic properties from other more easily obtained soil and land characteristics—are valuable tools (Padarian et al., 2018; Van Looy et al., 2017). PTFs that consider some soil structural variable can be particularly useful in modeling changes to soil hydraulic properties arising from alterations in soil structure (e.g., soil bulk density). Several studies have shown that including structural variables improves PTFs predictions. For example, in a study using 487 data points mined from the literature, Jorda et al. (2015) found that bulk density (ρ_b) and land use (a variable that most directly impacts soil structure) to be the most important predictors of hydraulic conductivity. Nguyen et al. (2015) found that prediction of soil moisture for their study sites was improved when soils were grouped by soil structural criteria. Other studies found that including soil structural variables in terms

of fractal parameters improved predictions of soil hydraulic properties (Huang & Zhang, 2005; Mohammadi et al., 2013).

The need for parameterization and inclusion of new soil structural variables in PTFs is widely recognized (Patil & Singh, 2016; Van Looy et al., 2017; Vereecken et al., 2010). However, soil structure remains poorly represented in PTFs; the lack of universally applicable and quantitative measures of soil structure remains to be a key challenge (Diaz-Zorita et al., 2002; Ghezzehei, 2011). The relationships between soil structural variables (such as bulk density, aggregate stability, aggregate size distribution, and organic matter concentration) and hydraulic properties are very complex and highly non-linear. It is extremely difficult to model these relationships accurately using physically-based models or traditional statistical methods. This opens an opportunity to revisit these challenges using data-driven methods which excel in such situations, such as machine learning (ML) techniques (Shen et al., 2018). This study was motivated by the growing availability of large databases of soil hydraulic properties and the current progress in ML tools.

The overarching aim of this work is to develop machine learning-based PTFs (ML-PTFs) for predicting K_s and advancing our quantitative knowledge of how soil structural indicators control hydraulic conductivity. The specific objectives of this study are to (1) develop a robust ML-PTFs, (2) identify important soil variables that control K_s, and (3) analyze the effect of soil structural alteration on K_s. Because of limitation on data availability, we used only bulk density (ρ_b) and total organic carbon content (C) as indicators of soil structure. These two variables are routinely characterized and reported in soil surveys and would make our models consistent with the objective of PTFs "to translate data we *have* to data we *need*" (Bouma, 1989).

2.2 Background

2.2.1 Indicators of Soil Structure

Indicators of soil structure that can have a direct or indirect effect on K_s are summarized in Table 2-1. Conceptual and mechanistic understanding of how several of these factors influence hydraulic conductivity has been the subject of numerous studies over the past decades. However, much of this knowledge remains qualitative or constrained to only a narrow range of soils. Although there has been considerable progress in linking topology and morphology of pore space (e.g., acquired via X-ray computed tomography) with hydraulic properties, the majority of this work involves advanced computation that is comparable to direct measurement in terms of the required effort. Perhaps the most glaring challenge is that only a few of these properties are characterized routinely. Moreover, generalizable quantitative indicators of soil structure that can be directly linked with hydraulic conductivity are very few. Of the listed parameters, ρ_b and organic matter are the two most common indicators of soil structure used in predicting hydraulic properties (e.g., Ary,a & Paris, 1981; Gupta & Larson, 1979; Nemes et al., 2005; Vereecken et al., 1989).

Table 2-1 Quantitative soil structure metrics and their significance to hydraulic conductivity.

Soil Variable	Significance and Mechanism	Example of Studies
Aggregation (size distribution, stability)	Indicates structure of macro- and meso-pores. Aggregate strength, type and compactness of aggregates	Koekkoek & Booltink (1999)
Bulk density	Indicates packing compaction. Influences total porosity, pore size distribution, and connectivity.	Schaapet al. (2001)
Clay type and metal oxides	Dominate properties that affect aggregation and water retention. Type and concentration of clay influence structure through aggregation, swell and shrink behavior, etc.	Rajkai & Varallyay (1992)
Fractal dimensions	Quantification of the heterogeneity, tortuosity and connectivity of soil pore/solid space.	Bayat et al (2013); Huang & Zhang (2005)
Mechanical properties and shrink-swell parameters (coefficient of linear extensibility, COLE)	Indicate dynamic properties of structure.	Baumer (1992); McKenzie et al. (1991); Watt et al. (1998)
Organic matter (Organic matter/carbon content)	Influences compaction, bulk density, aggregation, porosity and pore architecture.	Dexter et al. (2008); Nemes et al. (2005); Rawls et al. (2003)
Particle surface area	Indicates particle and pore sizes.	Bayat et al. (2013); Watt et al. (1998)
Penetration resistance	Indicates compactness and porosity.	Bayat & Ebrahim Zadeh (2018); Lipiec et al. (2009); Watt et al. (1998)
Porosity metrics (Mercury porosimetry or Imaging methods: Porosity, Connectivity, Pore-size distribution, Pore geometry.	Indicate actual pore architecture.	Otalvaro et al. (2016); Romero & Simms (2008)
Water retention characteristics (Characteristic water retention, retention curve fitting parameters, S-index)	Indicates the size distribution of pores.	Dexter (2004); Koekkoek & Booltink (1999); Rawls et al. (1982)

2.2.1.1 Bulk density

ρ_b is routinely measured during soil characterization and is used to estimate total porosity. Within a given textural class, variation in bulk density can be directly attributed to the degree of compactness (Hakansson & Lipiec, 2000) or aggregation (Aksakal et al., 2019). Therefore, ρ_b has been an essential variable in both physically based and empirical models of hydraulic conductivity (Assouline & Or, 2013 and references therein.). It is well

recognized that compaction of a given soil (hence, increase in ρ_b) leads to a reduction in saturated hydraulic conductivity (Assouline, 2006). However, the interactions of bulk density with other soil physical-chemical characteristics (such as texture and organic matter content) in influencing hydraulic conductivity is too complex to be captured by classical regression analyses or physically-based models. For example, in their study, Bogie et al. (2018) observed a significant decrease in both infiltration rate (a proxy of hydraulic conductivity) along with a decrease in ρ_b following a decade of organic matter input to a sandy soil.

2.2.1.2 Organic carbon content

Soil organic matter content is another routinely measured soil property that has a less direct, yet important, control on soil structure. Organic matter content affects soil structure largely because of its influence on soil aggregation, aggregate stability and associated porosity (Haynes & Beare, 1996; Hudson, 1994; Huntington, 2007). Generally, an increase in organic matter increases soil aggregate formation and aggregate stability and thus K_s (Hudson, 1994; Saxton & Rawls, 2006). Studies show that organic matter effects on soil hydraulic properties are similar to those of clay and high clay content reduces the effects of increased organic matter (Rawls et al., 2003; Saxton & Rawls, 2006). On the other hand, Dexter et al. (2008) found that organic matter substantially influences soil physical behavior (i.e., matrix and structural porosity) only when the clay content is above a threshold relative to C.

2.2.2 Machine Learning Algorithms

Numerous ML algorithms exist for multivariate regression modeling. In this study, we compared four popular ML algorithms: the k-nearest neighbors (KNN), support vector regression (SVR), random forest (RF), and boosted regression tree (BRT).

Many studies have used these ML algorithms in different problems related to soil hydraulic properties. Several studies have used KNN type of ML to predict soil hydraulic properties (e.g., Botula et al., 2013; Nemes et al., 2006, 2008); Elshorbagy et al. (2010) identify KNN as an attractive modeling technique for hydrology applications. Many studies have used SVR algorithm to model soil hydraulic properties (e.g., Angelaki et al., 2018; Kaingo et al., 2018; Kotlar et al., 2019; Mady & Shein, 2018; Singh et al., 2019). Recently, some studies have found that SVR models predicted soil hydraulic properties more accurately than artificial neural network models (Khlosi et al., 2016; Twarakavi et al., 2009; Zhang et al., 2018). Both the RF and BRT algorithms use an ensemble of regression trees as their base learners, and several studies have highlighted the power of these algorithms in predicting soil properties in general and hydraulic properties in particular. Hengl et al. (2017), for example, used RF and BRT among an ensemble of other models to build global soil map. Chaney et al. (2019) similarly employed RF to build a map of predicted soil properties over the United States. Recently Szabó et al. (2019) have developed PTFs based on RF and BRT to map soil hydraulic properties across a watershed. Koestel & Jorda (2014) showed that the RF algorithm can be used to accurately model soil preferential

solute transport. Jorda et al. (2015) used BRT models to predict K_s and explore important variables that control it.

2.2.2.1 K-nearest neighbors (KNN)

KNN are one of the simplest algorithms with respect to their underlying principle and, often, computational demand. Predictions for a new instance are made based on the average of the values of its 'k' nearest (i.e., most similar) neighbors in the training data. Nearest neighbors are commonly identified by Euclidean distances in the predictor parameter space. The number of 'k' nearest neighbors is the only parameter to tune during the training of KNN models.

2.2.2.2 Support vector regression (SVR)

SVR is an adaptation of the support vector machine (SVM) for regression problems (Cortes & Vapnik, 1995; Drucker et al., 1997). The SVM learning is a generalization of 'maximal margin classifier.' The algorithm first maps the input variables into a high-dimensional space using a fixed mapping function—a kernel function. The algorithm then constructs hyperplanes, which are used for classification or, in the case of SVR, for regression. In this study, we use the Radial Basis Function kernel which is one of the most commonly used kernels in SVR. Some advantages of SVR include the fact that they do not suffer from the problem of local minima, and that they have few parameters to tune when training the model.

2.2.2.3 Random forest (RF)

RF are popular models that are relatively simple to train and tune (Hastie et al., 2009). They apply ensemble techniques by averaging a large number of individual decision tree-based models. Tree models are 'grown' by searching for a predictor that ensures the best split that results in the smallest model error. The individual trees in RF ensemble are built on bootstrapped training sample and only a small group of predictor variables are considered at each split, this ensures that trees are de-correlated with each other (Breiman, 2001; James et al., 2013).

2.2.2.4 Boosted regression trees (BRT)

BRT, another form of decision tree model ensemble, enhances the model using the gradient boosting technique. The gradient boosting algorithm constructs additive regression models by sequentially fitting 'simple base learner' functions (i.e., decision trees) to current pseudo-residuals at each iteration (Friedman, 2002). These pseudo-residuals are the gradient of the loss function being minimized. BRT models have shown considerable success and often outperform other ML algorithms (Elith et al., 2008; Natekin & Knoll, 2013). BRT models are also particularly adept for less-than-clean data (Friedman, 2001), which makes them particularly attractive in our work where the training data is compiled from various sources and different measurement methods which makes it prone to some inconsistencies.

Tree-based models, both the RF and BRT, have the advantage of being able to rank predictor variables' relative importance. For a tree-based model, the approximate relative influence ($\hat{J}_j^2$) of a predictor variable x_j is calculated by Equation 2.1.

$$\hat{J}_j^2 = \sum_{splits\ on\ x_j} I_t^2 \tag{2.1}$$

where I_t^2 is the empirical improvement by splitting on predictor x_j at that point. For tree ensemble models, relative importance is given by averaging the relative influence of variable x_j across all trees of the model (Ridgeway, 2012).

2.3 Methods

The methods section is organized into four sections. The first section describes the training data and data pre-processing for the ML training. The second section describes the model training and testing procedures. The third section describes the predictor variable importance analysis procedure. The fourth section describes the methods used to test the response of K_s to perturbations on soil structural variables.

2.3.1 Data Preparation

The data used for training and testing PTFs is derived from the USKSAT database (Pachepsky & Park, 2015). The database contains K_s along with several textural, and structural information of over 27,000 US soils compiled from 45 datasets. For over 95 per cent of the soils, K_s was measured using a constant head method on samples sizes of approximately 5.5 cm length and 3 cm internal-diameter. In addition to the USKSAT database, we also acquired a subset of the USKSAT soils directly from Florida Soil Characterization Data, hereafter, FLSOIL (University of Florida, n.d.). The FLSOIL contains data of over 8,000 soils which are also part of the USKSAT but have additional variables of soil water contents at 11 pressure heads. We used the FLSOIL subset to build separate models that utilize water retention data in order to evaluate the effect of water retention variables in estimating K_s.

2.3.1.1 Data cleaning

From the USKSAT soils, we selected a subset with only 11 variables (see variables in Table 2-2). To prepare this subset for the ML procedure, we removed soils that had either a missing data in one or more of the variables or contained values that met one of our exclusion criteria shown in Equation 2.2a-d.

$$|1 - (\mathrm{Sa} + \mathrm{Si} + \mathrm{Cl})| > 0.05 \tag{2.2a}$$

$$|\mathrm{Sa} - (\mathrm{VCOS} + \mathrm{COS} + \mathrm{MS} + \mathrm{FS} + \mathrm{VFS})| > 0.05 \tag{2.2b}$$

$$\rho_b > 2.65 \tag{2.2c}$$

$$\theta(h_x) < \theta(h_y) \text{ for } h_x > h_y \tag{2.2d}$$

The exclusion criteria ensured that all soil texture fractions add up to 100 % (within a 5 % margin to account for possible significant digit and rounding issues). The criteria also ensure that there are no outlier bulk densities (Equation 2c) and, in the FLSOIL database, that water retention for soils do not increase with water tension. The resulting 'cleaned' USKSAT database contained 18,644 soils. Summary of the cleaned USKSAT database is shown in Table 2-2.

Table 2-2 Summary of cleaned USKSAT data variables.

	Variable (abbreviation)	Unit	Min.	Q1[a]	Median	Mean	Q3[a]	Max.
Measured	Saturated hydraulic conductivity (K_s)	$\log_e(cm\ hr^{-1})$	-7.5	0.68	2.6	1.9	3.4	6.7
	Bulk density (ρ_b)	$g\ cm^{-3}$	0.02	1.5	1.6	1.5	1.6	2.6
	Organic carbon content (C)	$\log_e(\%)$	-4.6	-2.3	-1.5	-1.4	-0.4	2.9
	Clay fraction (Cl)	%	0	1.3	3.1	8.7	12.7	93.4
	Silt fraction (Si)	%	0	2	3.8	5.6	6.5	94.5
	Sand fraction (Sa)	%	0.2	79.8	92.1	85.7	96.2	99.9
	Very coarse sand fraction (VCOS)	%	0	0	0	0.3	0.2	19.6
	Coarse sand fraction (COS)	%	0	0.6	2.1	3.8	5.1	44.6
	Medium sand fraction (MS)	%	0	7.6	16.8	20.9	30.1	77.7
	Fine sand fraction (FS)	%	0.1	35.0	49.6	50.3	65.6	97.4
	Very fine sand fraction (VFS)	%	0	3.8	8.2	10.4	14.8	56.4
Calculated	10th percentile particle size (d_{10})	µm	0.02	0.7	56.02	52.0	100.1	253.4
	50th percentile particle size (d_{50})	µm	0.17	133.9	156.4	167.1	193.3	534.1
	60th percentile particle size (d_{60})	µm	0.3	159.0	180.2	199.5	237.3	656.5
	Coefficient of Uniformity (CU)	$\log_e(-)$	0.47	1.0	1.2	2.8	5.33	8.7
	Complexed organic carbon (C_X)	%	0	0.1	0.1	0.2	0.3	5.19

[a] Q1 and Q3 are the first and third quartiles respectively.

From the FLSOIL database, we selected additional variables of volumetric water contents (θ) at 11 pressure heads (h), i.e. ,3.5 to 1500 cm H_2O. Measurements of K_s, ρ_b and θ in the FLSOIL database were made in replicates of either two or three. We used the arithmetic means of these variables. The resulting 'cleaned' FLSOIL database contained 5,985 soils.

The distribution of the cleaned USKSAT soils across the USDA textural classes and summary of K_s by textural classes is shown in Figure 2-1. While the textural composition of the cleaned USKSAT is dominated by sand which represents 63.4% of the database, 11 out of the 12 textural classes are represented.

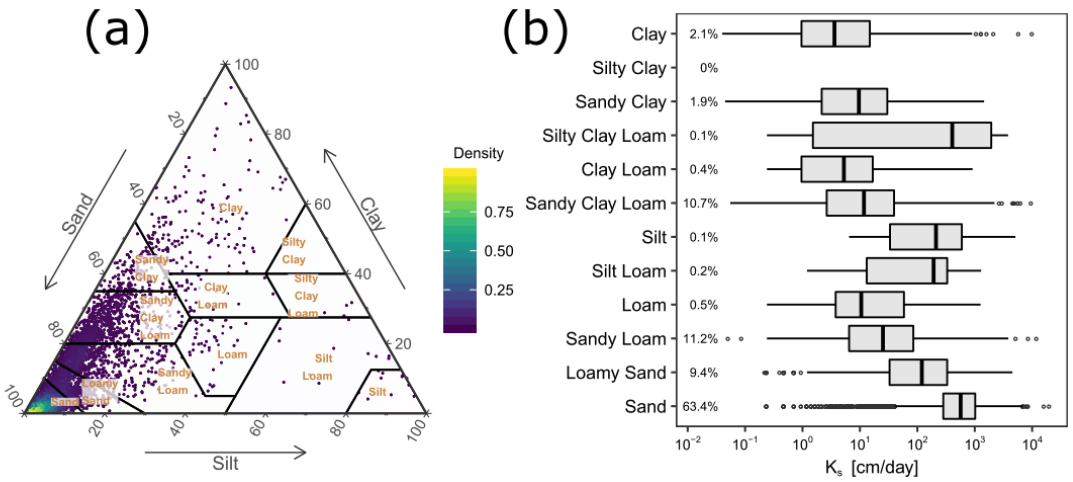

Figure 2-1 (a) Distribution of the cleaned USKSAT soils in USDA textural classes and (b) ranges of K_s and percent database by textural class.

2.3.1.2 Computed secondary soil variables

2.3.1.2.1 Particle size distribution

Particle size distribution has a strong influence on hydraulic conductivity and variables of particle size distribution, particularly 10th percentile particle size, have been used in several semiempirical models (e.g., Carrier (2003) and references within). We calculate the 10th, 50th and 60th percentile particle sizes (d_{10}, d_{50} and d_{60}, respectively) from soil textural fraction data. For this, we constructed cumulative particle size distribution by linear-interpolation of the seven texture sizes (2, 50, 100, 250, 500, 1000 and 2000 µm), and a very small diameter of 0.01 µm as the minimum size. We then calculate the d_{10}, d_{50} and d_{60} particle sizes from the fitted distribution. The coefficient of uniformity (CU) of the particle size distribution was calculated as $CU = d_{60}/d_{10}$ (Skaggs et al., 2001).

2.3.1.2.2 Complexed organic carbon

The concept of complexed organic carbon (C_X) was introduced by Dexter et al. (2008) to better describe the influence of organic matter on soil physical behavior. C_X is the proportion of C that forms complexes with the clay fraction and it is calculated with the assumption that 1 g of C is complexed with n g of clay mass. Thus, for sufficiently high clay content ($Cl > n\,C$) all the C can be complexed. C_X is computed as:

$$C_X = \min\left[C, \frac{Cl}{n}\right] \tag{3}$$

Dexter et al. (2008) found that $n = 10$ best described the physical behavior of their study soils where were from French and Polish databases. We used the same ratio to calculate C_X in our study.

2.3.2 Model Building

The overall procedure of building the ML models is illustrated in Figure 2-2. The computationally demanding steps of model training and testing were run using a high-performance computing cluster. The *caret* R package (Kuhn, 2017) was used to handle training and tuning procedures. The ML algorithms were implemented using the following R packages: KNN from the *kknn* package (Hechenbichler & Schliep, 2004), SVR from the *kernlab* package (Karatzoglou et al., 2004), RF from the *randomForest* package (Liaw & Wiener, 2015), and BRT from the *gbm* package (Ridgeway, 2017).

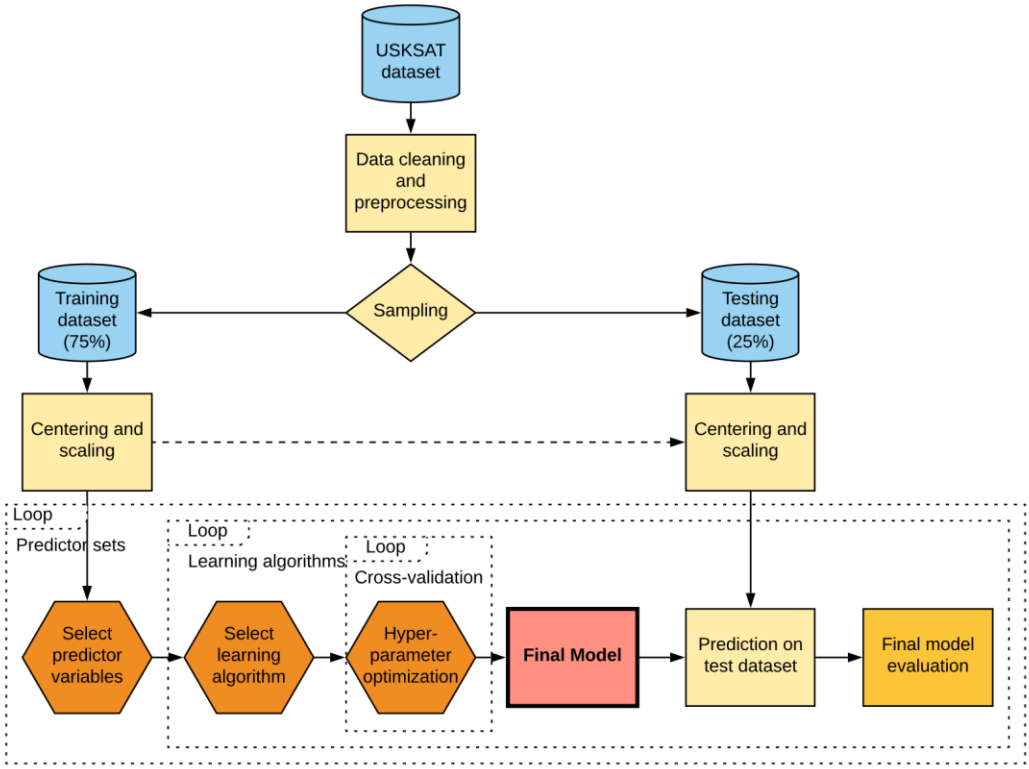

Figure 2-2 Flow chart of the model building process.

2.3.2.1 Data pre-processing

Data pre-processing prior to model training included the following. The K_s and C values were log transformed in order to make data more normally distributed (as inspected visually from density plot and Q-Q plot). Zero values of C were replaced with a small number of 0.001 prior to the log transformation. The USKSAT and FLSOIL databases were then split 75-25% into training and testing data sets, respectively. Prior to modeling, all variables excluding K_s were centered to the variable's mean and scaled by the variable's standard deviation in the training data (Equation 4):

$$x' = \frac{x - \bar{x}}{\sigma_x} \tag{4}$$

where x' is the centered and scaled value of variable x; $\bar{x}$ and σ_x are the respective arithmetic mean and standard deviation of the variable in the training dataset. The testing data sets are centered and scaled using the same mean and standard deviation values of the training set. No correlation ($R^2 > 0.8$) was detected amongst all possible pairs of predictor variables except d_{50}~d_{60} ($R^2 = 0.97$), and $Sand$~$Clay$ ($R^2 = 0.82$).

2.3.2.2 Predictor set hierarchy

It is desirable to select the minimum subset of the predictors needed to construct a model without a substantial reduction in prediction accuracy. The selection of such subset of predictors—feature selection—is done with the objectives of (a) improving prediction accuracy, (b) reducing model complexity, which makes interpretation of the effects of predictors easier giving us a better understanding of the underlying processes, and (c) reducing the amount of input variables needed to use the model.

In addition to testing multiple ML algorithms, we also built and analyzed multiple models with different sets of predictors. Throughout this manuscript, we refer to individual models by the ML algorithm used and the set of input predictors it takes, i.e., its predictor set hierarchy. We distinguish model hierarchy by appending a two-part numeric code separated by a hyphen where the first number denotes the number of textural variables the model uses (3, 7 or 10) and the second number denotes the number of structure related variables used (0, 1 or 2). The list of models by predictor set hierarchy is shown in Table 2-3. The lowest hierarchy model takes only the three textural size fractions. The three highest hierarchy models required variables of water retention and were trained on only the FLSOIL database. For these models only, we include a third numeric code which represents the number of water retention variables included (either 1, 2, or 11).

Table 2-3 Predictor set hierarchy codes and list of the input variables.

Hierarchy ID	Input Variables	Database
3-0	Cl, Si, Sa	USKSAT
3-1	Cl, Si, Sa, ρ_b	USKSAT
3-2	Cl, Si, Sa, ρ_b, C	USKSAT
7-0	Cl, Si, VFS, FS, MS, COS, VCOS	USKSAT
7-1	Cl, Si, VFS, FS, MS, COS, VCOS, ρ_b	USKSAT
7-2	Cl, Si, VFS, FS, MS, COS, VCOS, ρ_b, C	USKSAT
10-2	Cl, Si, VFS, FS, MS, COS, VCOS, d_{10}, d_{50}, CU, ρ_b, C	USKSAT
7-2-1	Cl, Si, VFS, FS, MS, COS, VCOS, ρ_b, C, $\theta(330)$	FLSOIL
7-2-2	Cl, Si, VFS, FS, MS, COS, VCOS, ρ_b, C, $\theta(330)$, $\theta(15000)$	FLSOIL
7-2-11	Cl, Si, VFS, FS, MS, COS, VCOS, ρ_b, C, $\theta(3.5)$, $\theta(20)$, $\theta(30)$, $\theta(45)$, $\theta(60)$, $\theta(80)$, $\theta(100)$, $\theta(150)$, $\theta(200)$, $\theta(330)$, $\theta(15000)$	FLSOIL

2.3.2.3 Model training

The selection of optimal model parameters in the model training process, i.e., model tuning, was done by *k*-fold cross-validation method. We used a five-times repeated, 10-fold cross validation method to select optimum model parameters using a comprehensive grid search method. Cross-validation is done to estimate the test error rate by holding out a subset of the training data (i.e., validation set) from the fitting process and then applying the fitted model to predict the validation subset. In *k*-fold cross-validation, the training data is randomly divided into *k* approximately equal subsets, and model fitting repeated *k* times; each time, treating a different subset as a validation set. This process allows the calculation of a validation set error rate which estimates the test error rate (James et al., 2013).

2.3.2.4 Model assessment

The final performance of models was assessed on the separate hold-out test dataset that was not used in the model training. The performance of models is measured in terms of root mean squared log-transformed error (RMSLE), mean log-transformed error (MLE) and the coefficient of determination (R^2) determined as follows:

$$RMSLE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\log_{10}(\widehat{K_s})_i - \log_{10}(K_s)_i\right]^2} \tag{5}$$

$$MLE = \frac{1}{N}\sum_{i=1}^{N}\left[\log_{10}(\widehat{K_s})_i - \log_{10}(K_s)_i\right] \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left[\log_{10}(\widehat{K_s})_i - \log_{10}(K_s)_i\right]^2}{\sum_{i=1}^{N}[\log_{10}(\overline{K_s})_i - \log_{10}(K_s)_i]^2} \tag{7}$$

where N is the number of observations; K_s is the measured value; $\widehat{K_s}$ is the predicted value; and $\overline{K_s}$ is the mean of measured values.

The RMSLE indicates the average deviation of predictions from the measured value with smaller values indicating better performance. The MLE measures systematic bias, positive or negative values indicate the average tendency of the predicted values to be larger or smaller than the measured values, respectively. The R^2 indicates correspondence between predicted and measured data with higher values indicating stronger correspondence.

2.3.2.5 Comparison with other PTFs

Further evaluation of our models was done by comparing our models with ten other PTF models frequently cited in the literature (Abdelbaki et al., 2009; Ghanbarian et al., 2017). The required predictors for each of the alternative models tested are given in Table 2-4. The Nemes et al. (2005) model is the only one of the alternative models that use organic matter content. We converted C to organic matter as: $OM = C \times 1.724$. This conventional conversion ratio is not accurate on all soils (Pribyl, 2010), but we considered it good

enough for our purpose after not observing meaningful differences in the overall performance by using different ratios that range between1.5 and 2. Porosity values required by the Brakensiek et al. (1984) and Saxton et al. (1986) models were calculated from ρ_b as: $\phi = 1 - (\rho_b/2.65)$. The Rosetta-3 model was run using the Python code made available by Zhang and Schaap (2017); we implemented the rest of models in R. The R code to calculate all the alternative PTFs we tested in this study is available online at https://github.com/saraya209/soil_ksat.

Table 2-4 Some proposed PTFs frequently used in literature to estimate K_s and their required input variable.

Reference – Model Name	Required Input Variables
Zhang & Schaap (2017) – Rosetta-3 (H3)	Cl, Si, Sa, ρ_b
Ghanbarian et al. (2015) – SHC2	Cl, Si, Sa, ρ_b, L, D[a]
Nemes et al. (2005)	Cl, Sa, ρ_b, C, OM[b]
Campbell & Shiozawa (1994)	Cl, Si
Dane & Puckett (1994)	Cl
Jabro (1992)	Cl, Si, ρ_b
Saxton et al. (1986)	Cl, Sa, ϕ
Puckett et al. (1985)	Cl
Brakensiek et al. (1984)	Cl, Sa, ϕ[c]
Cosby et al. (1984)	Cl, Si

[a] L = sample length (cm) and D = sample internal-diameter (cm); [b] OM = organic matter content (%); [c] ϕ = porosity (-).

2.3.3 Predictor Variable Importance

The predictor variable importance is the statistical significance of each predictor variable with respect to its effect on the generated model. For the tree-based models, RF and BRT, variable importance is calculated internally within the model algorithm (Equation 1). For the rest of the ML models, we calculated the predictor variable importance by recursive feature elimination method which is done by recursively removing predictors before training a model and evaluating the change in model performance. In this method, to account for possible bias in variable subset selection (Ambroise & McLachlan, 2002; Hastie et al., 2009), we included a separate layer of 10-fold cross-validation to the entire sequence modeling steps.

2.3.4 Sensitivity of K_s to Structural Perturbations

We used the ML-PTFs developed in this study to test the sensitivity of K_s to perturbations of soil structural variables (i.e. ρ_b and C). Sensitivity of K_s was analyzed by using our best performing model (BRT-7-2) to predict the marginal effect of varying one of the variables while keeping the others constant (Hastie et al., 2009; Hochachka et al., 2007; James et al., 2013). Interactive effect of ρ_b and C perturbations on K_s sensitivity was similarly analyzed by varying both structural variables together. Because the effects of ρ_b and C on

morphology and topology of soils are likely to be dependent on soil texture, the partial dependence relationships were analyzed separately for each textural class.

To construct partial dependence relationships, we randomly selected 100 soils, with replacement, from each textural class of the cleaned USKSAT dataset. The BRT-7-2 model was used to predict K_s of each sample while incrementally perturbing one of the two structural variables. To analyze the effect of ρ_b, the ρ_b value of each soil was incrementally varied from 0.5 to 2 g cm^{-3} while keeping the other variables of that soil constant.

Partial dependence of K_s with changes of C was determined similarly by incrementally varying C of each soil from 0.03 to 10 %. When perturbing the values of C, however, we also changed the values of ρ_b according to a linear correlation equation we developed between ρ_b and $\log(C)$ for each textural class based on USKSAT dataset (Appendix A, Figure A0-1). We did this to account for the observed relationship between C and ρ_b. For each incremental change of C value, we changed the value of ρ_b by a normal random variate about the mean and variance of the linear correlation fit.

We also analyzed the interactive effect of both structural variables on K_s by perturbing both ρ_b and C simultaneously for each of the 100 sampled soils. In this analysis, ρ_b and C were varied independently.

2.4 Results and Discussion

2.4.1 Model Performances

The model performance analyses are organized by the ML algorithm and the predictor set hierarchies.

The model performances, in terms of RMSLE, for all the learning algorithms and predictor set hierarchies combination are shown in Figure 2-3. These performance tests were conducted on the test set which includes 4,661 soils (25% of the cleaned USKSAT dataset). The BRT algorithm consistently outperformed the other learning algorithms across all predictor set hierarchies with the RF algorithm closely behind. Performance for all ML algorithms generally increased with an increase in the number of predictors used. The one exception was the performance of KNN algorithm which decreased when ρ_b was included from KNN-3-0 to KNN-3-1. Including sand subclass fractions led to large improvement on model performances across all learning algorithms. Using the seven textural size fractions instead of only three improved performance by a larger proportion than that of including ρ_b or C variables with only three texture size fractions.

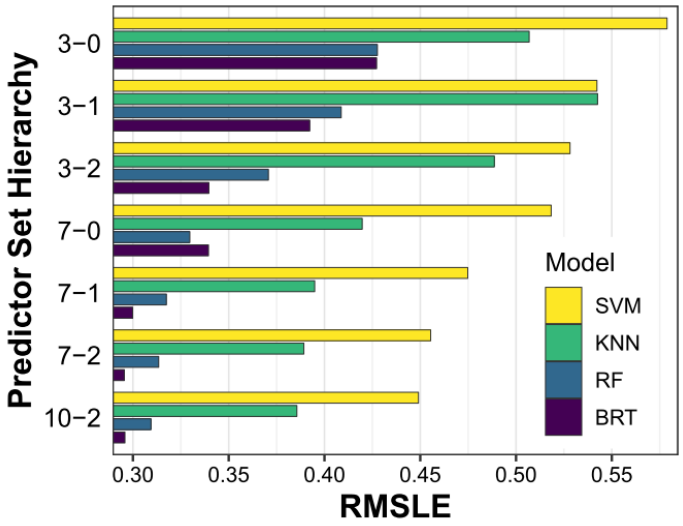

Figure 2-3 Model performance in terms of RMSLE by machine learning algorithm type and number of predictor variables used (see Table 2-3). BRT-7-2 and 10-2 are the best models with lowest prediction error.

2.4.2 Comparison with Other PTFs

One-to-one scatter plot comparison of four of our models with the alternative models we tested is shown in Figure 2-4. The predicted and measured values are assigned to the x-axis and y-axis, respectively, as recommended by Piñeiro et al. (2008). Our models outperformed all ten alternatives. The revised version of the popular PTF, Rosetta-3 (Zhang & Schaap, 2017), showed the best performance among all alternative models we tested. Based on the MLE and the 1 to 1 plot, the Rosetta-3 model slightly tends to overestimate K_s, particularly at lower magnitudes. The BRT-3-1 model is equivalent to the Rosetta-3 model we tested in terms of the required predictor variables to run the model (i.e. Sa, Si, Cl and ρ_b). In terms of RMSLE and MLE, as well as the data clouds on the 1:1 line, all hierarchies of our model showed good performance across the full range of the K_s values. The distribution of the residuals for all hierarchies of our models appears similar, with increased performance (tighter distribution around the 1:1 line) for larger magnitudes of K_s.

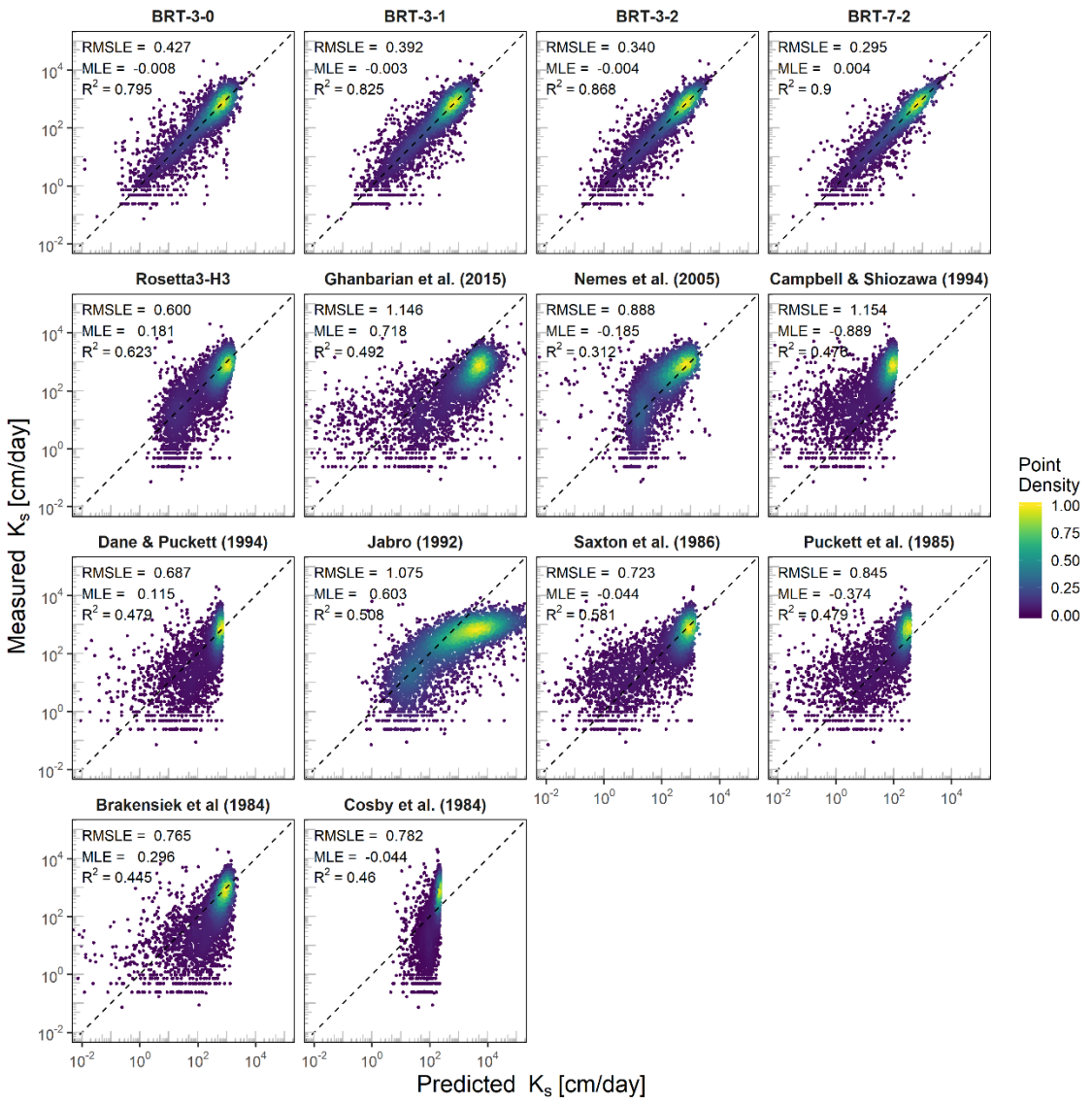

Figure 2-4 Comparisons between measured and model predicted saturated hydraulic conductivities for the testing data set (n = 4,661). Top row panels show 1 to 1 comparison of predictions made using four different input hierarchy BRT models we developed. The remaining ten panels show predictions made using other commonly used PTF models. Note that the number of samples is 4,540 for the evaluation of Jabro (1992) model (samples with either zero silt or clay content were removed); 4,562 for Nemes et al. (2005) model (samples with 0 cm/day prediction were removed); and 4, 650 for the evaluation of Rosetta-3 model (samples with *NA* prediction were removed). The color scale denotes density of the points estimated by 2-D kernel density estimation of the values.

The distribution of residuals of the best performing BRT-7-2 model is shown in Figure 2-5.

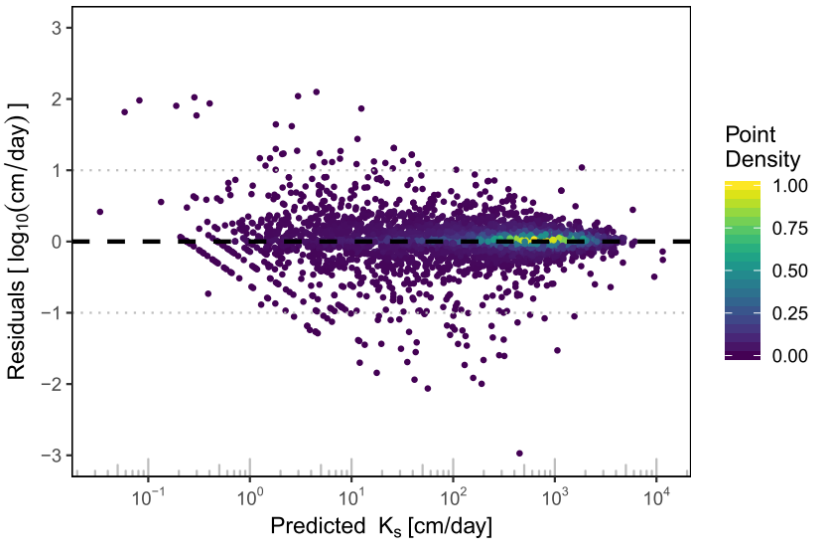

Figure 2-5 Residuals versus predicted values for the best performing BRT-7-2 model (n = 4,661).

The spread of the residuals is larger at the smaller magnitude predictions. Most of the predictions fall within an order of magnitude of the measured values and only very few predictions are off by more than two orders of magnitude. The performance within each textural class also showed that all hierarchies of our models performed better than the alternative models we tested. Figure 2-6 shows the performances of four selected hierarchies of our model and the Rosetta-3 within the soil textural classes.

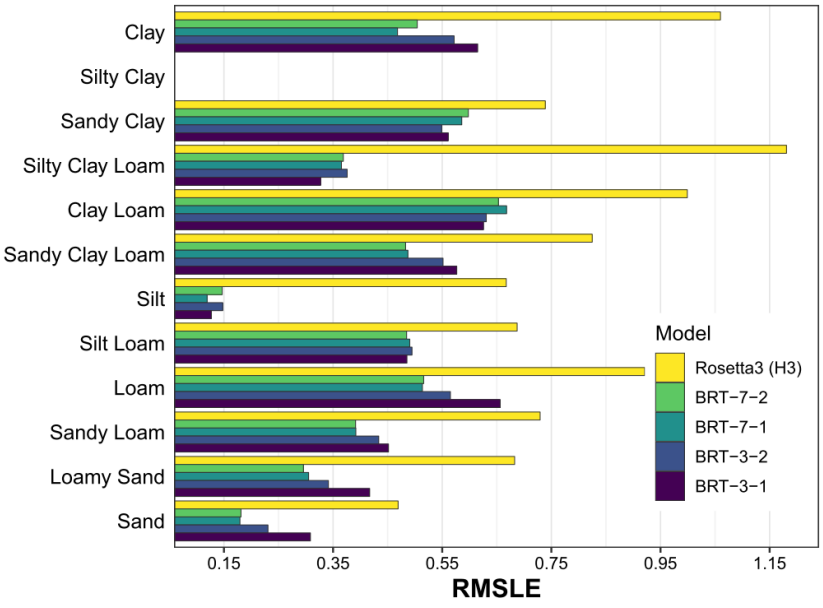

Figure 2-6 Model performances on USKSAT test dataset by textural class.

Performance statistics of our best model (BRT-7-2) by textural classes is given in Table 2-5. The RMSLE for the model ranged from 0.147 for silt group to 0.653 for clay loam. R^2 ranged from the 0.996 for silt clay loam group to 0.497 for sandy clays.

Table 2-5 Best model (BRT-7-2) performance metrics by textural classes. Bold numbers indicate the largest and smallest values within each metric.

Texture Class	Count of soils	RMSLE	R^2	MLE
Clay	83	0.504	0.634	0.028
Silty Clay	0	-	-	-
Sandy Clay	89	0.598	**0.497**	-0.044
Silty Clay Loam	3	0.369	**0.996**	-0.184
Clay Loam	25	**0.653**	0.529	**-0.247**
Sandy Clay Loam	532	0.483	0.673	0.025
Silt	6	**0.147**	0.968	-0.076
Silt Loam	9	0.485	0.741	0.208
Loam	20	0.516	0.732	0.049
Sandy Loam	500	0.391	0.763	0.017
Loamy Sand	443	0.296	0.844	0.001
Sand	2951	0.181	0.875	**0.0001**
Overall	**4661**	**0.295**	**0.900**	**0.004**

2.4.3 Predictor Variable Importance

Both best performing ML algorithms, RF and BRT, showed a similar ranking of variable importance. The relative variable importance ranking for the best performing models are shown in Figure 2-7.

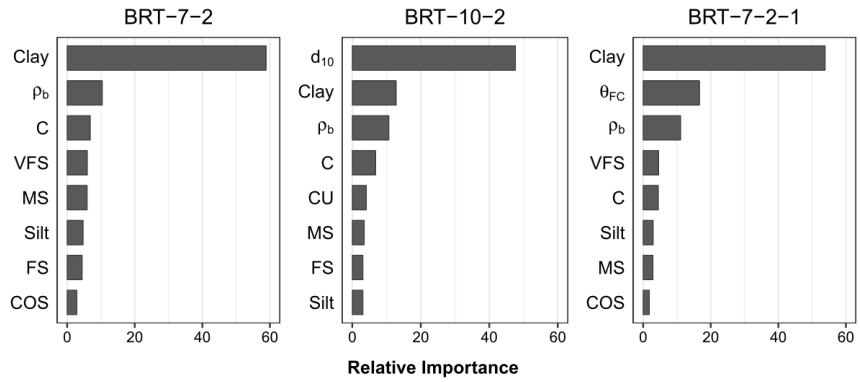

Figure 2-7 Relative importance ranking of the top eight predictors for three different hierarchy models.

For the best performing model (BRT-7-2), the most important predictor was clay mass present followed by ρ_b and C. The dominance of the clay content as the most important variable even though 73% of the training data was classified as sand or loamy sand, with <15% clay highlights the disproportionate importance of the fine particles to K_s. It is also notable that the two structural indicators (ρ_b and C) were ranked as the second and third most important variables. When the variable d_{10} is included (BRT-10-2) it overtakes clay and becomes overwhelmingly the most important predictor. However, including d_{10} did not improve the model performance in terms of RMSLE, which suggests the ML algorithm is able to 'learn' the importance of d_{10} from the raw textural size data where the d_{10} parameter is calculated from. Including C_X did not lead to model improvement and the $C_X$variable was ranked as the least important. Models that include C_X variable are hence not included in this paper.

2.4.4 Importance of Water Retention Variables

The analysis of water retention variable was done using models trained on only the FLSOIL database, a much smaller database but which had water retention values. Models trained on only the FLSOIL database had lower performance than those trained using the entire USKSAT database. In order to compare variable importance of water retention on relative bases, we trained similar hierarchy models on FLSOIL database. When looking at variable importance, water retention at field capacity (θ_{FC}) was the second most important variable, preceded only by clay content (Figure 2-7). The addition of a single water retention variable (i.e. θ_{FC}) led to a relatively large improvement in model performance with an RMSLE drop of 13% from RMSLE of 0.49 to 0.42 (Figure 2-8). Water content at field capacity is a strong indicator of soil structure and its importance in predicting K_s in our models highlights the importance of structure in K_s.

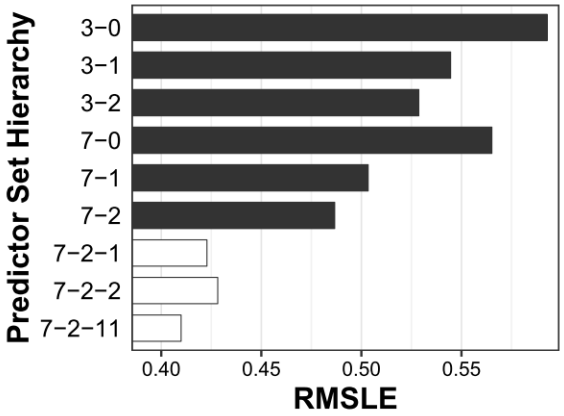

Figure 2-8 Performance of BRT models trained with FLSOIL database only. White bars represent models that include water retention variables (see Table 2-3).

2.4.5 Prediction Interval

Providing uncertainty estimates in PTF predictions is important to assess the reliability of estimates (Schaap & Leij, 1998). Uncertainty estimates are also essential information in most applications such as use in land surface models (Baroni et al., 2017; Chaney et al., 2019; Folberth et al., 2016; Van Looy et al., 2017). Prediction intervals can be estimated by building an ensemble of models. The RF algorithm is an ensemble of regression trees, and prediction intervals can easily be calculated from the variance of the ensemble trees. Although the BRT models slightly outperformed the RF models, the possibility of producing a prediction interval may make these models more appealing choice in circumstances where knowledge of the prediction uncertainty is essential, such as long-term trends in soil processes using land surface models. **Figure 2-9** demonstrates the prediction intervals from the best performing RF model (RF-7-2). The figure shows histograms of the deviations from measured values of 500 individual tree predictions for each texture for the 100 randomly selected soils within each texture group. The 75^{th} percentile prediction intervals for almost all soils fall within an order of magnitude of the mean prediction.

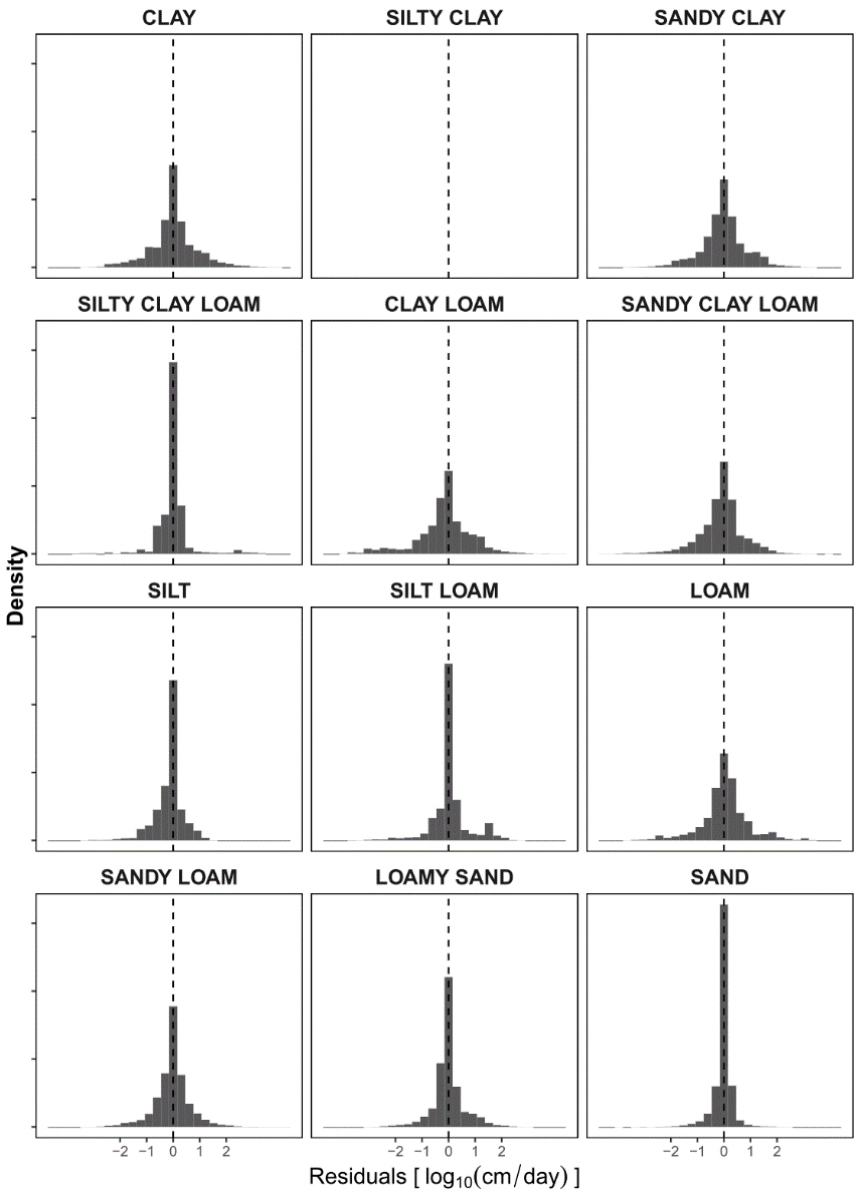

Figure 2-9 Histogram of K_s prediction deviations from the measured values for a subset of 500 individual regression trees that make up the RF-7-2 model. Prediction are done on 100 soils selected randomly with replacement from each textural class of the USKSAT test dataset.

2.4.6 Sensitivity of K_s to Structural Perturbations

2.4.6.1 Bulk density

Figure 2-10 shows the change of K_s prediction with ρ_b. K_s decreased with an increase in ρ_b and a more uniform pattern is apparent when soils are grouped by textural calss. The K_s pattern of change appears to follow an inverted s-curve. To enable a more quantitative description of the sensitivity, we fitted a $K_s \sim \rho_b$ logistic-curve within each texture (Equation 8)

$$K_s' = \frac{1}{1 + e^{k_\rho(\rho_b - \rho_{b.0})}} \tag{8}$$

where $\rho_{b.0}$ is the mid-point (point of inflection) of the curve and k_ρ is the slope of the curve at $\rho_b = \rho_{b.0}$. On average across all textural classes the maximum change in K_s occurs at $\rho_{b.0} \approx 1.8\ \mathrm{g\ cm^{-3}}$. The steepness of the curve averaged at ≈ -2, and ranged from $k_\rho = -1.1$ for silty loam soils to steepest rate of $k_\rho = -13.28$ for sandy clay soils. The trend of silty clay loam was not well approximated by logistic curve having its inflection point outside the range, this is likely because the number of records in this class were too few to generalize to a trend for the texture class.

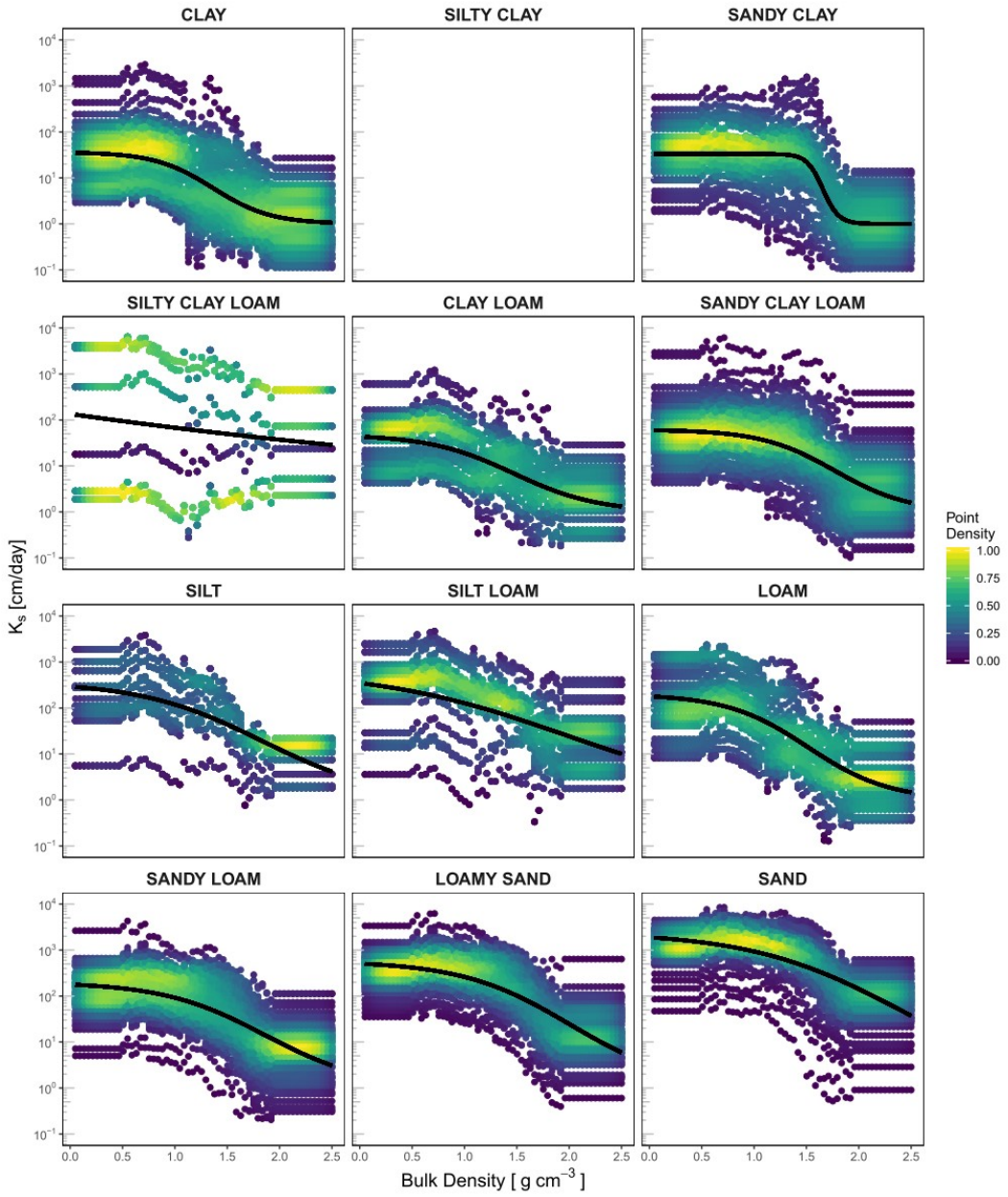

Figure 2-10 Predicted K_s changes across ρ_b for 100 randomly selected soils from soil textural groups. Black trend lines show logistic curve fit.

2.4.6.2 Organic carbon content

The relation between C content and K_s was difficult to discern and only became apparent when the relationships were plotted separately for each soil textural class (Figure 2-11). Nemes et al. (2005) have also noted this lack of generalizable explanatory trend between C and K_s .

The K_s across all the textural classes except for the two coarsest classes—loamy sand and sand—increased with increase in C content. This trend is consistent with C being associated

with structure development (e.g., aggregation and formation of biopores and macropores) which increases the overall permeability of soils. For sand and loamy sand soils, the apparent slow down and even reversal of the trends at higher C contents (Figure 2-11) may be due to the inherently high proportion of large pores in soils of this textural groups. This trend may also suggest that an increase in C content is associated with the shrinking of larger pores in these coarse textured soils by increased aggregation. Similar effect of C in reducing K_s of sandy soils while increasing that of finer textured soils was observed by Nemes et al. (2004, 2005); for soils that are 50% sand and clay content between ~25 to 45%, they reported lower K_s prediction for those soils with higher C ($C = 5\%$ compared to those with $1\% \leq C \leq 3\%$) which led them to conclude that the relationship between C and K_s is very complex.

To enable quantitative analysis of the sensitivity, we fitted $K_s \sim C$ logistic-curves to each texture class individually as:

$$K_s' = \frac{1}{1 + e^{k_C(C - C_0)}} \tag{9}$$

where C_0 is the mid-point (point of inflection) of the curve and k_C is the slope of the curve at $C = C_0$. The trends of K_s change with C generally appears weaker than that of with ρ_b. For the loamy sand and sand soil groups, however, the relationship is notably different with a decrease in K_s at $C \geq 3\%$.

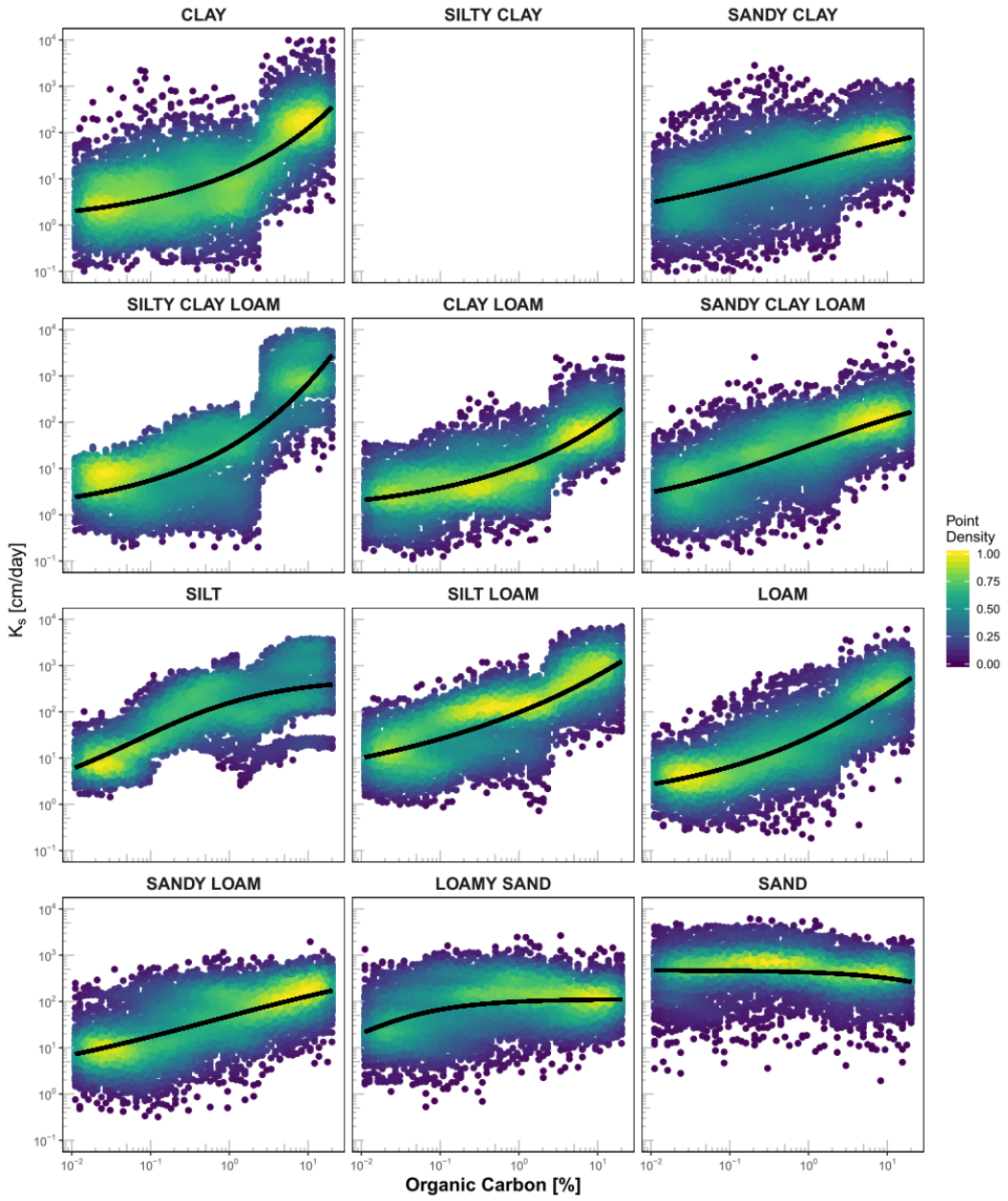

Figure 2-11 Predicted K_s changes across C for 100 randomly selected soils from soil textural groups. Black trend lines are logistic curve fits.

To visualize the combined effect of ρ_b and C changes on K_s, we plotted a two-dimensional heatmap of K_s shown in Figure 2-12. The heatmap shows normalized K_s predictions across ρ_b and C changes of 100 randomly selected soils for each textural class. The predicted K_s were normalized to range from 0 to 1 for each soil using Equation 10.

$$K_s' = \frac{K_s - K_s^{\min}}{K_s^{\max} - K_s^{\min}} \tag{10}$$

where K_s' is the normalized K_s, and $K_s^{\min}$ and $K_s^{\max}$ are the minimum and maximum K_s values of the soil.

For all the classes, except loamy sand and sand, the highest and lowest K_s are in the top-left and bottom-right corners, respectively. The trends were significantly altered for the coarse textures. The lowest K_s are in the top-right corner. These observations suggest that the effects of C on soil structure and ultimately K_s are not masked by changes in ρ_b. The combined effect of C and ρ_b on K_s is illustrated in **Figure 2-12**.

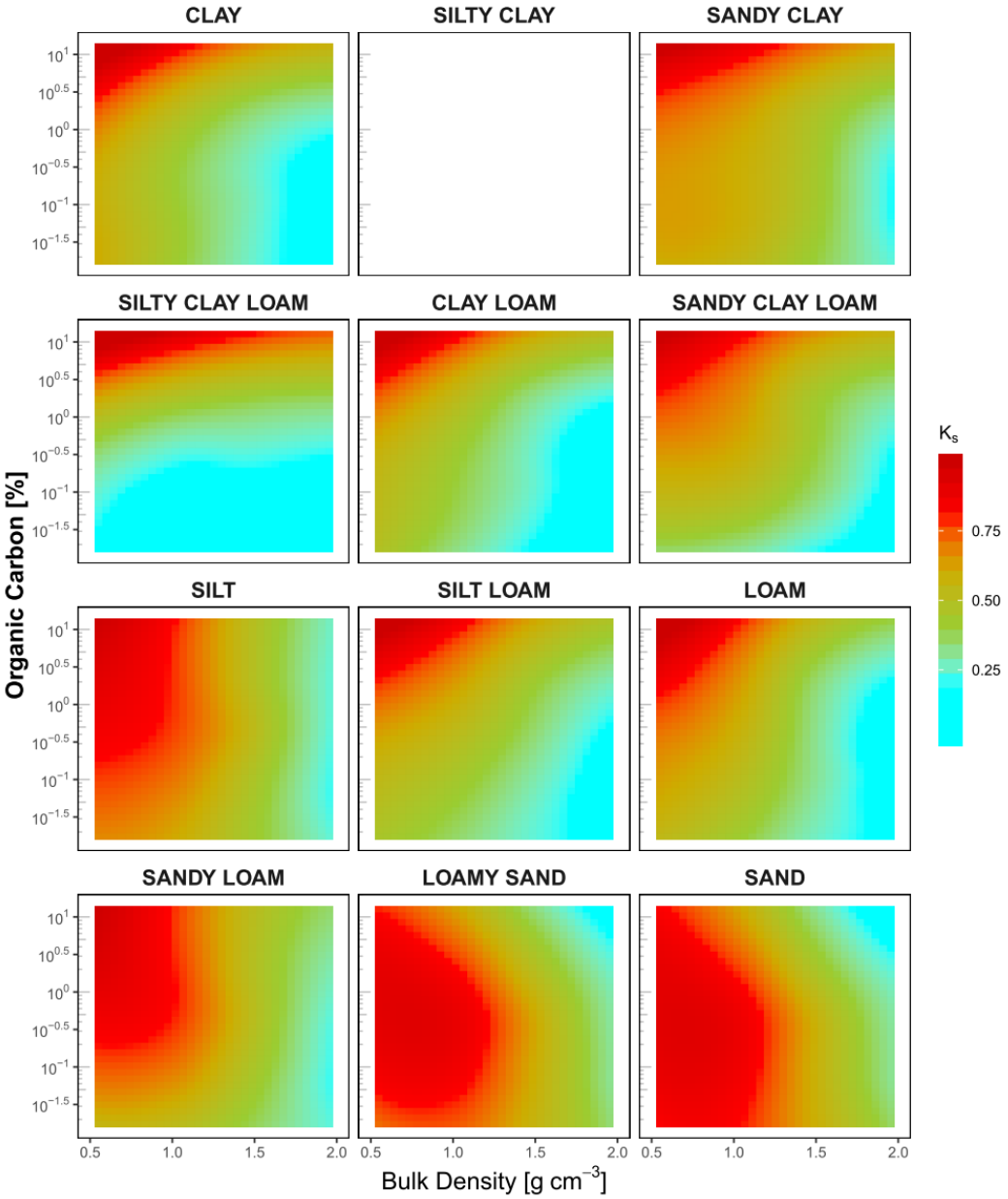

Figure 2-12 Heatmap of predicted K_s values (scaled $\log_e(cm/hr)$) across ρ_b and C changes for 100 randomly selected soils from soil textural groups. K_s values in the heatmap have been smoothed with LOESS for clarity of display.

2.5 Conclusions

We developed ML-based PTFs that were trained on large soil database (USKSAT). We tested four popular ML algorithms (KNN, SVR, RF and BRT) with a range of predictor set hierarchies. The BRT models outperformed the other ML algorithms closely followed by the RF models. The best performing BRT model has prediction accuracy for $\log_{10} K_S$ of $RMSE = 0.295$. This RMSLE is lower by 50% than the reported accuracy for the revised version of the popular Rosetta-3 model, $RMSLE = 0.6$, (Zhang & Schaap, 2017). The accuracy achieved by our models in this study to predict K_s is far higher than any other PTF model we are aware of.

Based on their relative importance to predict K_s, d_{10} was by far the most important predictor for K_s followed by clay content. However, removing d_{10} did not reduce model performance, suggesting that the algorithms were able to learn the effect of d_{10} from the raw textural data. Following these two textural variables, the most important variables were ρ_b and C content which highlights the importance of these structural variables on K_s. The potential impact of structural perturbations on K_s was illustrated by the functional relationships between the structural variables and K_s. We observed that the effects of structural perturbation on K_s varied with textural classes. Generally, K_s decreased with an increase in ρ_b, with maximum sensitivity at $\rho_b \approx 1.8$ g cm^{-3}. The effect of C perturbation on K_s was more complex. For all textural classes, except loamy sand and sand, increasing C led to an increase in K_s. Whereas for the coarsest textural class, increasing C reduced K_s. These trends may suggest that C induced aggregation increases the relative proportion of large pores (inter-aggregate pores) in fine and medium textured soils. Whereas in sandy soils, aggregation increases the proportion of fine intra-aggregate pores. These functional relationships demonstrate the potential that such models can be incorporated in land-surface and soil-systems models and be used to account for the sensitivity of infiltration and water flow to soil structural alterations and disturbances (e.g., organic matter accumulation or tillage).

Data and Code Availability

The saturated hydraulic conductivity pedotransfer models are publicly available at https://doi.org/10.6071/M3T95H. The raw and cleaned version of the datasets used (USKSAT and FLSOIL), and the R scripts used to generate and analyze models are available online at https://github.com/saraya209/soil_ksat.

To enhance the usability of the models developed, we developed an application with graphical user interface (Figure 2-13). The application can be used to predict saturated hydraulic conductivity for a single soil or multiple soils in a table. The application is freely available to the public and can be run locally in any machine with RStudio software installed.

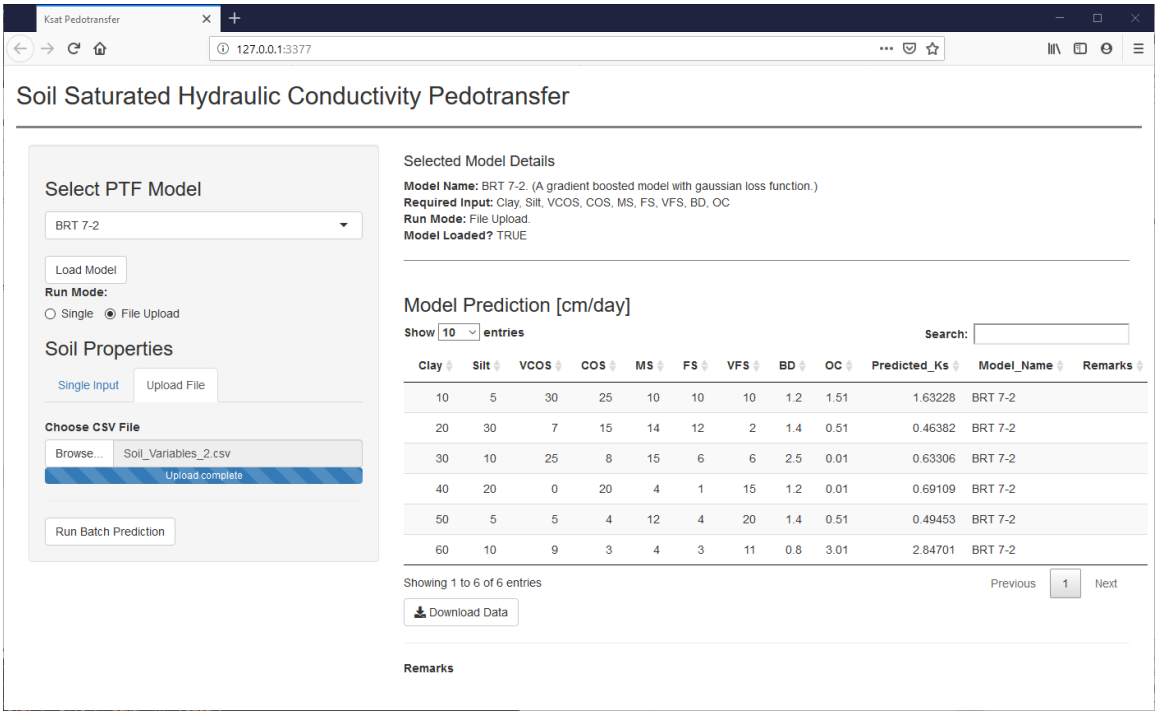

Figure 2-13 The graphical user interface for the pedotransfer app developed.

Acknowledgement

We thank Dr. Yakov A Pachepsky (Environmental Microbial and Food Safety Lab, USDA) for making the USKSAT data available to us. We also thank Dr. Attila Nemes (Norwegian Institute for Bioeconomy Research) for his thorough review of the manuscript and insightful comments. We gratefully acknowledge computing time on the Multi-Environment Computer for Exploration and Discovery (MERCED) cluster at UC Merced, which was funded by National Science Foundation Grant No. ACI-1429783.

References

Abdelbaki, A. M., Youssef, M. A., Naguib, E. M. F., Kiwan, M. E., & El-giddawy, E. I. (2009). Evaluation of Pedotransfer Functions for Predicting Saturated Hydraulic Conductivity for U.S. Soils. *2009 Reno, Nevada, June 21 - June 24, 2009*. https://doi.org/10.13031/2013.27309

Aksakal, E. L., Barik, K., Angin, I., Sari, S., & Islam, K. R. (2019). Spatio-temporal variability in physical properties of different textured soils under similar management and semi-arid climatic conditions. *Catena*, *172*(January 2018), 528–546. https://doi.org/10.1016/j.catena.2018.09.017

Alaoui, A., Lipiec, J., & Gerke, H. H. (2011). A review of the changes in the soil pore system due to soil deformation: A hydrodynamic perspective. *Soil and Tillage Research*, *115–116*, 1–15. https://doi.org/10.1016/j.still.2011.06.002

Ambroise, C., & McLachlan, G. J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences*, *99*(10), 6562–6566. https://doi.org/10.1073/pnas.102102699

Angelaki, A., Singh Nain, S., Singh, V., & Sihag, P. (2018). Estimation of models for cumulative infiltration of soil using machine learning methods. *ISH Journal of Hydraulic Engineering*, *00*(00), 1–8. https://doi.org/10.1080/09715010.2018.1531274

Arya, L. M., & Paris, J. F. (1981). A Physicoempirical Model to Predict the Soil Moisture Characteristic from Particle-Size Distribution and Bulk Density Data. *Soil Science Society of America Journal*, *45*(6), 1023. https://doi.org/10.2136/sssaj1981.03615995004500060004x

Assouline, S. (2006). Modeling the Relationship between Soil Bulk Density and the Hydraulic Conductivity Function Contribution of the Agricultural Research Organization, Institute of Soil, Water and Environmental Sciences, Bet Dagan, Israel, No. 608/05. *Vadose Zone Journal*, *5*, 697–705. https://doi.org/10.2136/vzj2005.0084

Assouline, S., & Or, D. (2013). Conceptual and Parametric Representation of Soil Hydraulic Properties: A Review. *Vadose Zone Journal*, *12*(4), 0. https://doi.org/10.2136/vzj2013.07.0121

Baroni, G., Zink, M., Kumar, R., Samaniego, L., & Attinger, S. (2017). Effects of uncertainty in soil properties on simulated hydrological states and fluxes at different spatio-temporal scales. *Hydrology and Earth System Sciences*, *21*(5), 2301–2320. https://doi.org/10.5194/hess-21-2301-2017

Baumer, O. W. (1992). Predicting unsaturated hydraulic parameters. In M. T. van Genuchten, F. J. Leij, & L. J. Lund (Eds.), *Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of*

Unsaturated Soils. (pp. 341–354). Riverside, CA: University of California, Riverside.

Bayat, H., & Ebrahim Zadeh, G. (2018). Estimation of the soil water retention curve using penetration resistance curve models. *Computers and Electronics in Agriculture*, *144*(April 2017), 329–343. https://doi.org/10.1016/j.compag.2017.10.015

Bayat, H., Neyshaburi, M. R., Mohammadi, K., Nariman-Zadeh, N., & Irannejad, M. (2013). Improving water content estimations using penetration resistance and principal component analysis. *Soil and Tillage Research*, *129*, 83–92. https://doi.org/10.1016/j.still.2013.01.009

Beven, K., & Germann, P. (1982). Macropores and water flow in soils. *Water Resources Research*, *18*(5), 1311–1325. https://doi.org/10.1029/WR018i005p01311

Bittelli, M., Campbell, G. S., & Tomei, F. (2015). *Soil Physics with Python*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199683093.001.0001

Bogie, N. A., Bayala, R., Diedhiou, I., Conklin, M. H., Fogel, M. L., Dick, R. P., & Ghezzehei, T. A. (2018). Hydraulic Redistribution by Native Sahelian Shrubs: Bioirrigation to Resist In-Season Drought. *Frontiers in Environmental Science*, *6*(September), 1–12. https://doi.org/10.3389/fenvs.2018.00098

Botula, Y.-D., Nemes, A., Mafuka, P., Van Ranst, E., & Cornelis, W. M. (2013). Prediction of Water Retention of Soils from the Humid Tropics by the Nonparametric -Nearest Neighbor Approach. *Vadose Zone Journal*, *12*(2), 1-17. https://doi.org/10.2136/vzj2012.0123

Bouma, J. (1989). Using Soil Survey Data for Quantitative Land Evaluation. In B. A. Stewart (Ed.), *Advances in Soil Science* (pp. 177–213). New York, NY: Springer US. https://doi.org/10.1007/978-1-4612-3532-3_4

Brakensiek, D. L., Rawls, W. J., & Stephesnson, G. R. (1984). Modifying SCS hydrologic soil groups and curve numbers for rangeland soils. *ASAE Paper No. PNR-84-203.*

Breiman, L. (2001). Random Forest. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brooks, E. S., Boll, J., & McDaniel, P. A. (2004). A hillslope-scale experiment to measure lateral saturated hydraulic conductivity. *Water Resources Research*, *40*(4), 1–10. https://doi.org/10.1029/2003WR002858

Campbell, G. S., & Shiozawa, S. (1994). Prediction of hydraulic properties of soils using particle-size distribution and bulk density data. In M. T. van Genuchten, F. J. Leij, & L. J. Lund (Eds.), *Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils* (pp. 317–328).

Berkeley, CA, USA: University of California Press.

Carrier, W. D. (2003). Goodbye, Hazen; Hello, Kozeny-Carman. *Journal of Geotechnical and Geoenvironmental Engineering*, *129*(November), 1054–1056. https://doi.org/10.1061/(ASCE)1090-0241(2003):11(1054)

Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., … Yimam, Y. (2019). POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties Over the Contiguous United States. *Water Resources Research*, *55*(4), 2916–2938. https://doi.org/10.1029/2018WR022797

Chivenge, P. P., Murwira, H. K., Giller, K. E., Mapfumo, P., & Six, J. (2007). Long-term impact of reduced tillage and residue management on soil carbon stabilization: Implications for conservation agriculture on contrasting soils. *Soil & Tillage Research*, *94*(2), 328–337. https://doi.org/DOI 10.1016/j.still.2006.08.006

Cortes, C., & Vapnik, V. N. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Cosby, B. J., Hornberger, G. M., Clapp, R. B., & Ginn, T. R. (1984). A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resources Research*, *20*(6), 682–690. https://doi.org/10.1029/WR020i006p00682

Dane, J. H., & Puckett, W. (1994). Field soil hydraulic properties based on physical and mineralogical information. In *Proceedings of the international workshop on indirect methods for estimating the hydraulic properties of unsaturated soils. University of California, Riverside* (pp. 389–403).

de Almeida, W. S., Panachuki, E., de Oliveira, P. T. S., da Silva Menezes, R., Sobrinho, T. A., & de Carvalho, D. F. (2018). Effect of soil tillage and vegetal cover on soil water infiltration. *Soil and Tillage Research*, *175*(June 2017), 130–138. https://doi.org/10.1016/j.still.2017.07.009

Dexter, A. R. (2004). Soil physical quality: Part III: Unsaturated hydraulic conductivity and general conclusions about S-theory. *Geoderma*, *120*(3), 227–239. https://doi.org/10.1016/j.geoderma.2003.09.006

Dexter, A. R., Richard, G., Arrouays, D., Czyz, E. A., Jolivet, C., & Duval, O. (2008). Complexed organic matter controls soil physical properties. *Geoderma*, *144*(3–4), 620–627. https://doi.org/10.1016/j.geoderma.2008.01.022

Diaz-Zorita, M., Perfect, E., & Grove, J. H. (2002). Disruptive methods for assessing soil structure. *Soil and Tillage Research*, *64*(1–2), 3–22. https://doi.org/10.1016/S0167-1987(01)00254-9

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. N. (1997). Support vector regression machines. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.),

Advances in Neural Information Processing Systems 9 (Vol. 1, pp. 155–161). MIT Press. Retrieved from http://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elshorbagy, A., Corzo, G., Srinivasulu, S., & Solomatine, D. P. (2010). Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. *Hydrology and Earth System Sciences*, *14*(10), 1943–1961. https://doi.org/10.5194/hess-14-1943-2010

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., & van der Velde, M. (2016). Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nature Communications*, *7*(1), 11872. https://doi.org/10.1038/ncomms11872

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1017/CBO9781107415324.004

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Ghanbarian, B., Taslimitehrani, V., Dong, G., & Pachepsky, Y. A. (2015). Sample dimensions effect on prediction of soil water retention curve and saturated hydraulic conductivity. *Journal of Hydrology*, *528*, 127–137. https://doi.org/10.1016/j.jhydrol.2015.06.024

Ghanbarian, B., Taslimitehrani, V., & Pachepsky, Y. A. (2017). Accuracy of sample dimension-dependent pedotransfer functions in estimation of soil saturated hydraulic conductivity. *CATENA*, *149*, 374–380. https://doi.org/10.1016/j.catena.2016.10.015

Ghezzehei, T. A. (2011). Soil Structure. In P. M. Huang, Y. Li, & M. E. . Sumner (Eds.), *Handbook of Soil Sciences: Properties and Processes* (pp. 1–18). Boca Raton, FL, USA: CRC Press.

Ghezzehei, T. A., & Or, D. (2000). Dynamics of soil aggregate coalescence governed by capillary and rheological processes. *Water Resources Research*, *36*(2), 367–379. https://doi.org/10.1029/1999WR900316

Gupta, S. C., & Larson, W. E. (1979). Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. *Water Resources Research*, *15*(6). https://doi.org/10.1029/WR015i006p01633

Hakansson, I., & Lipiec, J. (2000). A review of the usefulness of relative bulk density values in studies of soil structure and compaction. *Soil and Tillage Research*, *53*(2),

71–85. https://doi.org/10.1016/S0167-1987(99)00095-1

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). New York, NY: Springer New York. https://doi.org/10.1007/b94608

Haynes, R. J., & Beare, M. H. (1996). Aggregation and Organic Matter Storage in Mesothermal, Humid Soils. In M. R. Carter & B. A. Stewart (Eds.), *Structure and Soil Organic Matter Storage in Agricultural Soils* (pp. 213–262). Boca Raton, FL: CRC Press.

Hechenbichler, K., & Schliep, K. P. (2004). Weighted k-Nearest-Neighbor Techniques and Ordinal Classification. Ludwig-Maximilians University Munich. Retrieved from http://epub.ub.uni-muenchen.de/

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., … Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hochachka, W. M., Caruana, R., Fink, D., Munson, A., Riedewald, M., Sorokina, D., & Kelling, S. (2007). Data-Mining Discovery of Pattern and Process in Ecological Systems. *Journal of Wildlife Management*, *71*(7), 2427. https://doi.org/10.2193/2006-503

Huang, G., & Zhang, R. (2005). Evaluation of soil water retention curve with the pore–solid fractal model. *Geoderma*, *127*(1–2), 52–61. https://doi.org/10.1016/j.geoderma.2004.11.016

Hudson, B. D. (1994). Soil organic matter and available water capacity. *Journal of Soil and Water Conservation*, *49*(2), 189–194. Retrieved from http://www.jswconline.org/content/49/2/189.abstract

Huntington, T. G. (2007). Available Water Capacity and Soil Organic Matter. In *Encyclopedia of Soil Science* (2nd ed., pp. 139–143). New York: Taylor and Francis. https://doi.org/10.1081/E-ESS-120018496

Jabro, J. D. (1992). Estimation of Saturated Hydraulic Conductivity of Soils From Particle Size Distribution and Bulk Density Data. *Am. Soc. Agric. Eng.*, *35*, 557–560.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Jorda, H., Bechtold, M., Jarvis, N., & Koestel, J. (2015). Using boosted regression trees to explore key factors controlling saturated and near-saturated hydraulic conductivity. *European Journal of Soil Science*, *66*(4), 744–756. https://doi.org/10.1111/ejss.12249

Kaingo, J., Tumbo, S. D., Kihupi, N. I., & Mbilinyi, B. P. (2018). Prediction of Soil Moisture-Holding Capacity with Support Vector Machines in Dry Subhumid Tropics. *Applied and Environmental Soil Science*, *2018*, 1–10. https://doi.org/10.1155/2018/9263296

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab: An S4 Package for Kernel Methods in R (version 0.9-25). *Journal of Statistical Software*, *11*(9), 1–20. Retrieved from http://www.jstatsoft.org/v11/i09/

Khlosi, M., Alhamdoosh, M., Douaik, A., Gabriels, D., & Cornelis, W. M. (2016). Enhanced pedotransfer functions with support vector machines to predict water retention of calcareous soil. *European Journal of Soil Science*, *67*(3), 276–284. https://doi.org/10.1111/ejss.12345

Koekkoek, E. J. W., & Booltink, H. (1999). Neural network models to predict soil water retention. *European Journal of Soil Science*, *50*(3), 489–495. https://doi.org/10.1046/j.1365-2389.1999.00247.x

Koestel, J., & Jorda, H. (2014). What determines the strength of preferential transport in undisturbed soil under steady-state flow? *Geoderma*, *217–218*, 144–160. https://doi.org/10.1016/j.geoderma.2013.11.009

Kotlar, A. M., Iversen, B. V., & de Jong van Lier, Q. (2019). Evaluation of Parametric and Nonparametric Machine-Learning Techniques for Prediction of Saturated and Near-Saturated Hydraulic Conductivity. *Vadose Zone Journal*, *18*(1). https://doi.org/10.2136/vzj2018.07.0141

Kuhn, M. (2017). The caret Package. Retrieved from http://topepo.github.io/caret/index.html

Kuncoro, P. H., Koga, K., Satta, N., & Muto, Y. (2014). A study on the effect of compaction on transport properties of soil gas and water I: Relative gas diffusivity, air permeability, and saturated hydraulic conductivity. *Soil and Tillage Research*, *143*, 172–179. https://doi.org/10.1016/j.still.2014.02.006

Liaw, A., & Wiener, M. (2015). Package "randomForest" (v.4.6-12). *R Package*. https://doi.org/10.5244/C.22.54

Lipiec, J., Wójciga, A., & Horn, R. (2009). Hydraulic properties of soil aggregates as influenced by compaction. *Soil and Tillage Research*, *103*(1), 170–177. https://doi.org/10.1016/j.still.2008.10.021

Mady, A. Y., & Shein, E. V. (2018). Support Vector Machine and Nonlinear Regression Methods for Estimating Saturated Hydraulic Conductivity. *Moscow University Soil Science Bulletin*, *73*(3), 129–133. https://doi.org/10.3103/S0147687418030079

McKenzie, N. J., Smetten, K. R. J., & Ringrose-Voase, A. J. (1991). Evaluation of methods for inferring air and water propertiesof soils from feld morphology.

Australian Journal of Soil Research, *29*, 587–602.

Mohammadi, K., Bayat, H., Gregory, A., Neyshaburi, M., Irannejad, M., & Nariman-Zadeh, N. (2013). Combination of artificial neural networks and fractal theory to predict soil water retention curve. *Computers and Electronics in Agriculture*, *92*, 92–103. https://doi.org/10.1016/j.compag.2013.01.005

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*(DEC), 1–11. https://doi.org/10.3389/fnbot.2013.00021

Nemes, A., Rawls, W. J., & Pachepsky, Y. A. (2004). Effect of soil organic carbon on soil hydraulic properties. In Y. A. Pachepsky & W. J. Rawls (Eds.), *Developments in Soil Science* (Vol. 30, pp. 95–114). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S0166-2481(04)30006-1

Nemes, A., Rawls, W. J., & Pachepsky, Y. A. (2005). Influence of Organic Matter on the Estimation of Saturated Hydraulic Conductivity. *Soil Science Society of America Journal*, *69*(4), 1330. https://doi.org/10.2136/sssaj2004.0055

Nemes, A., Rawls, W. J., & Pachepsky, Y. A. (2006). Use of the Nonparametric Nearest Neighbor Approach to Estimate Soil Hydraulic Properties. *Soil Science Society of America Journal*, *70*(2), 327. https://doi.org/10.2136/sssaj2005.0128

Nemes, A., Roberts, R. T., Rawls, W. J., Pachepsky, Y. A., & van Genuchten, M. T. (2008). Software to estimate −33 and −1500kPa soil water retention using the non-parametric k-Nearest Neighbor technique. *Environmental Modelling and Software*, *23*(2), 254–255. https://doi.org/10.1016/j.envsoft.2007.05.018

Nguyen, P. M., De Pue, J., Le, K. Van, & Cornelis, W. (2015). Impact of regression methods on improved effects of soil structure on soil water retention estimates. *Journal of Hydrology*, *525*, 598–606. https://doi.org/10.1016/j.jhydrol.2015.04.014

Nielsen, J. E., Karup, D., de Jonge, L. W., Ahm, M., Bentzen, T. R., Rasmussen, M. R., & Moldrup, P. (2018). Can the Volume Ratio of Coarse to Fine Particles Explain the Hydraulic Properties of Sandy Soil? *Soil Science Society of America Journal*, *82*(5), 1093. https://doi.org/10.2136/sssaj2018.02.0083

Or, D., Leij, F. J., Snyder, V., & Ghezzehei, T. A. (2000). Stochastic model for posttilage soil pore space evolution. *Water Resources Research*, *36*(7), 1641–1652. https://doi.org/10.1142/S0217732309028163

Otalvaro, I. F., Neto, M. P. C., Delage, P., & Caicedo, B. (2016). Relationship between soil structure and water retention properties in a residual compacted soil. *Engineering Geology*, *205*, 73–80. https://doi.org/10.1016/j.enggeo.2016.02.016

Pachepsky, Y. A., & Park, Y. (2015). Saturated Hydraulic Conductivity of US Soils Grouped According to Textural Class and Bulk Density. *Soil Science Society of America Journal*, *79*(4), 1094. https://doi.org/10.2136/sssaj2015.02.0067

Padarian, J., Morris, J., Minasny, B., & McBratney, A. B. (2018). Pedotransfer Functions and Soil Inference Systems. In A. McBratney, B. Minasny, & U. Stockmann (Eds.), *Pedometrics. Progress in Soil Science.* (pp. 195–220). Springer, Cham. https://doi.org/10.1007/978-3-319-63439-5_7

Patil, N. G., & Singh, S. K. (2016). Pedotransfer Functions for Estimating Soil Hydraulic Properties: A Review. *Pedosphere*, *26*(4), 417–430. https://doi.org/10.1016/S1002-0160(15)60054-6

Piñeiro, G., Perelman, S., Guerschman, J. P., & Paruelo, J. M. (2008). How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, *216*(3–4), 316–322. https://doi.org/10.1016/j.ecolmodel.2008.05.006

Pribyl, D. W. (2010). A critical review of the conventional SOC to SOM conversion factor. *Geoderma*, *156*(3–4), 75–83. https://doi.org/10.1016/j.geoderma.2010.02.003

Puckett, W. E., Dane, J. H., & Hajek, B. F. (1985). Physical and Mineralogical Data to Determine Soil Hydraulic Properties. *Soil Science Society of America Journal*, *49*, 831–836.

Rajkai, K., & Varallyay, G. (1992). Estimating soil water retention from simpler properties by regression techniques. In M. T. van Genuchten, F. J. Leij, & L. J. Lund (Eds.), *Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils* (pp. 417–426). Riverside, CA: University of California, Riverside.

Rawls, W. J., Brakensiek, D. L., & Saxton, K. E. (1982). Estimation of Soil Water Properties. *Transactions of the ASAE*, *25*(5), 1316–1320. https://doi.org/10.13031/2013.33720

Rawls, W. J., Pachepsky, Y. A., Ritchie, J. C., Sobecki, T. M., & Bloodworth, H. (2003). Effect of soil organic carbon on soil water retention. *Geoderma*, *116*(1–2), 61–76. https://doi.org/10.1016/S0016-7061(03)00094-6

Ridgeway, G. (2012). *Generalized Boosted Models: A guide to the gbm package*. *Compute*. https://doi.org/10.1111/j.1467-9752.1996.tb00390.x

Ridgeway, G. (2017). Package “gbm.”

Romero, E., & Simms, P. H. (2008). Microstructure Investigation in Unsaturated Soils: A Review with Special Attention to Contribution of Mercury Intrusion Porosimetry and Environmental Scanning Electron Microscopy. *Geotechnical and Geological Engineering*, *26*(6), 705–727. https://doi.org/10.1007/s10706-008-9204-5

Saxton, K., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal*, *70*, 1569–1578. https://doi.org/10.2136/sssaj2005.0117

Saxton, K., Rawls, W. J., Romberger, J., & Papendick, R. (1986). Estimating generalized soil-water characteristics from texture. *Soil Science Society of America Journal*, *50*, 1031–1036. https://doi.org/10.1017/CBO9781107415324.004

Schaap, M. G., & Leij, F. J. (1998). Database-related accuracy and uncertainty of pedotransfer functions. *Soil Science*, *163*(10), 765–779. https://doi.org/10.1097/00010694-199810000-00001

Schaap, M. G., Leij, F. J., & van Genuchten, M. T. (2001). Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, *251*(3–4), 163–176. https://doi.org/10.1016/S0022-1694(01)00466-8

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F. J., … Tsai, W. P. (2018). HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. *Hydrology and Earth System Sciences*, *22*(11), 5639–5656. https://doi.org/10.5194/hess-22-5639-2018

Singh, B., Sihag, P., Pandhiani, S. M., Debnath, S., & Gautam, S. (2019). Estimation of permeability of soil using easy measured soil parameters: assessing the artificial intelligence-based models. *ISH Journal of Hydraulic Engineering*, *00*(00), 1–11. https://doi.org/10.1080/09715010.2019.1574615

Skaggs, T. H., Arya, L. M., Shouse, P. J., & Mohanty, B. P. (2001). Estimating Particle-Size Distribution from Limited Soil Texture Data. *Soil Science Society of America Journal*, *65*, 1038. https://doi.org/10.2136/sssaj2001.6541038x

Soil Science Division Staff. (2017). *Soil Survey Manual*. (C. Ditzler, K. Scheffe, & H. C. Monger, Eds.), *USDA, Natural Resources Conservation Service*. Washington, DC: Government Printing Office. Retrieved from https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054262

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., & Pásztor, L. (2019). Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. *Hydrology and Earth System Sciences*, *23*(6), 2615–2635. https://doi.org/10.5194/hess-23-2615-2019

Twarakavi, N. K. C., Simunek, J., & Schaap, M. G. (2009). Development of Pedotransfer Functions for Estimation of Soil Hydraulic Parameters using Support Vector Machines. *Soil Science Society of America Journal*, *73*(5), 1443. https://doi.org/10.2136/sssaj2008.0021

University of Florida. (n.d.). Florida Soil Characterization Database. *Department of Soil and Water Science*. Retrieved from http://soils.ifas.ufl.edu/flsoils/databaseintro.asp

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., … Vereecken, H. (2017). Pedotransfer Functions in Earth System Science: Challenges and Perspectives. *Reviews of Geophysics*, *55*(4), 1199–1256.

https://doi.org/10.1002/2017RG000581

Vereecken, H., Maes, J., Feyen, J., & Darius, P. (1989). Estimating the Soil Moisture Retention Characteristic From Texture, Bulk Density, and Carbon Content. *Soil Science*, *148*(6), 389–403. Retrieved from https://insights.ovid.com/crossref?an=00010694-198912000-00001

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y. A., Schaap, M. G., & van Genuchten, M. T. (2010). Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil Hydraulic Properties: A Review. *Vadose Zone Journal*, *9*(4), 795. https://doi.org/10.2136/vzj2010.0045

Watt, J. P. C., Pachepsky, Y. A., Rawls, W. J., & Gimene, D. (1998). Use of soil penetration resistance and group method of data handling to improve soil water retention estimates. *Soil & Tillage Research*, *49*, 117–126.

Zhang, Y., & Schaap, M. G. (2017). Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*, *547*, 39–53. https://doi.org/10.1016/j.jhydrol.2017.01.004

Zhang, Y., Schaap, M. G., & Zha, Y. (2018). A High-Resolution Global Map of Soil Hydraulic Properties Produced by a Hierarchical Parameterization of a Physically Based Water Retention Model. *Water Resources Research*, *54*(12), 9774–9790. https://doi.org/10.1029/2018WR023539

Chapter 3 Long-Term Impact of Cover Crop and Reduced Disturbance Tillage on Soil Pore Size and Soil Water Storage

Abstract

The effect of 18 years of the conservation agriculture as applied through reduced disturbance tillage (RT) and cover cropping (CC), on soil structural and hydraulic properties was investigated for a clay loam soil in the Central Valley of California. In addition to soil bulk density, porosity, and soil hydraulic functions, soil water storage and retention were evaluated using numerical HYDRUS simulations. Changes in pore size distribution (PSD) and conductivity associated with RT and CC improved soil structure and increased infiltration rate and soil water retention at unsaturation. Both RT and CC reduced soil water content at field capacity (-33 kPa suction) and plant available water content as determined in the laboratory. However, numerical simulations of irrigated tomatoes across the irrigated season showed that RT and CC potentially increased soil profile water storage and water availability. This result highlights the weakness of using static laboratory-measured properties to evaluate soil hydrology.

List of Acronyms and Symbols

CC	Cover crop
HCF	Hydraulic conductivity function
PAW	Plant available water content
PSD	Pore size distribution
RT	Reduced disturbance tillage
ST	Standard tillage
WRC	Water retention curve

K_s	Saturated hydraulic conductivity
θ_{FC}	Volumetric water content at field capacity
θ_{PWP}	Volumetric water content at permanent wilting point (-15 MPa suction)
ρ_b	Bulk density
ρ_p	Particle density
h	negative water suction ($h = -\psi$)
K	Hydraulic conductivity
θ	Volumetric water content
ψ	Matric potential
ϕ	Total porosity

3.1 Introduction

Conservation agriculture practices reduce negative impacts associated with conventional soil management (Mitchell et al., 2019; others). Negative effects associated with standard (conventional) tillage (ST) practices include degradation of soil structure, erosion, loss of nutrients, and reduction in soil microbial diversity and soil organic matter (Lal et al., 2007; Zuber & Villamil, 2016). Conservation agriculture is typically characterized by one, or a combination of three linked principles: (1) reduced mechanical soil disturbance, (2) preservation of a permanent organic soil cover, and (3) diversification of crop species (Kassam et al., 2019; Li et al., 2018; Mitchell et al., 2019). The adoption of conservation agriculture has been growing worldwide at an increasing rate since the 1960s. Between 2008 and 2015, the global area under conservation agriculture grew by 69% to 180 M ha (Blanco-Canqui & Ruis, 2018; Kassam et al., 2019). The adoption of conservation agriculture practices continues to grow rapidly in Californian's Central Valley. Mitchell et al. (2016) reported that the cultivated area under conservation agriculture for tomato and corn production in the Central Valley has increased from less than 5,000 ha in 2004 to over 140,000 ha in 2012.

Reduced disturbance tillage systems have been shown to sequester carbon and decrease greenhouse has emission (Palm et al., 2014; Reicosky & Allmaras, 2003; Sanz-Cobena et al., 2017); increase soil fertility (Veenstra et al., 2007, 2006); increase microbial biomass, richness and activity (Johnson & Hoyt, 1999; Martens, 2004; Zuber & Villamil, 2016); and improve environmental quality (Baker et al., 2005; Madden et al., 2008; Reicosky & Allmaras, 2003) without compromising yield (Alvarez & Steinbach, 2009; Naab et al., 2017; Rasmussen, 1999) while reducing cost (González-Sánchez et al., 2016; Mitchell, Klonsky, et al., 2009; Upadhyaya et al., 2001). Concerns with reduced tillage systems are that it may lead to soil consolidation and increase compaction. There are some contrasting finding on the effect of reduced disturbance tillage on soil physical health (Blanco-Canqui & Ruis, 2018).

Cover cropping, defined as planting between cropping seasons to maintain soil coverage throughout the year, is often practiced along with reduced or zero-tillage. Cover crops (CC) maintain soil cover, generate residues, and add diversity in conservation agriculture systems (Mitchell et al., 2019). Cover crops (CC) have been shown to reduce erosion (Reicosky & Forcella, 1998; Shelton et al., 2000), increase soil macroporosity (Abdollahi et al., 2014; Burr-Hersey et al., 2017), increase fertility and water storage (Ashworth et al., 2017; A. D. Basche, Archontoulis, et al., 2016; Duchene et al., 2017; Finney et al., 2017; Liu et al., 2019), increase microbial biomass, richness and activity(Duchene et al., 2017; Fernandez et al., 2016) , and reduce disease and pest pressure (Mitchell et al., 2017). Concerns with the practice of cover cropping include issues of difficulty during seeding activities and the possibility that the CC may harbor seedling pests and rodents (Mitchell et al., 2017).

In terms of the effect of conservation agriculture practices on soil hydrology, CC has been shown to improve soil water storage and infiltration (A. Basche & DeLonge, 2017;

Hudson, 1994; Johnson & Hoyt, 1999). Mitchell (2017)found that soil infiltration increased in reduced disturbance tillage plots with CC treatments and that CC treatment increased infiltration 2.8 times compared to ST soils without CC. Based on a meta-analysis from 27 studies, Basche and DeLonge (2017) conclude that CC was effective in enhancing soil water storage and other soil hydrologic properties when practiced for longer-term (> 10 years) and in drier environments (< 9000 mm annual rainfall).

Reduced disturbance tillage practices are sometimes referred to as conservation tillage which is a broad term that represents tillage practices that have a conservation goal (Mitchell, Pettygrove, et al., 2009; Reicosky & Allmaras, 2003). The California Conservation Agricultural Systems Innovation Center defines conservation tillage as a range of production practices that reduce primary intercrop tillage operations and either preserve 30% or more residue cover or reduce the total number of tillage passes by 40% or more (Mitchell, 2016). Throughout this manuscript, we will use the more descriptive RT instead of conservation tillage.

Several authors have noted the critical lack of field studies and the need for evaluation of long-term effects of conservation agriculture on the soil physical and hydraulic properties and soil hydrological processes (Bacq-Labreuil et al., 2019; A. Basche & DeLonge, 2017; Blanco-Canqui & Ruis, 2018; Peña-Sancho et al., 2016). This study aims to assess the impacts of long-term RT and CC practices on soil pore size distribution and related soil physical and hydraulic properties, and their potential for improving water infiltration and water storage. We hypothesize that RT and CC practices in an arid irrigated agricultural system will improve soil structure, thus positively affecting the soil's hydraulic properties and soil water storage and plant availability.

3.2 Methods

To measure the long-term impact of RT and CC practices on soil structure, we measured the soil bulk density (ρ_b), porosity (ϕ), pore size distribution (PSD), and soil hydraulic properties (water retention curve, WRC, and hydraulic conductivity function, HCF) of a field soils that have been under continuous RT with and without CC (RT-CC and RT-NO, respectively), and ST with and without CC (ST-CC and ST-NO, respectively) management for 18-years. To measure the impact of soil hydraulic property changes on profile water retention and storage, we conducted numerical irrigation simulations and evaluated the soil hydrology in terms of water infiltration, profile redistribution, and retention across an irrigated season.

3.2.1 Site

The study site was located at the University of California West Side Research and Extension Center in Five Points, California (Figure 3-1). The experimental plot has two-factor replicated treatments of tillage and winter cover cropping: standard tillage with and without cover crops (ST-NO and ST-CC, respectively); and reduced disturbance tillage with and without cover crops (RT-NO and RT-CC, respectively). Each treatment

combination was replicated eight times in a randomized complete block implemented on a 9 by 82-m dimension plot with an approximately 10-m buffer guard between the tillage treatments. All tractor and implement traffic were restricted to the furrows and planting beds were never moved. While the operations used varied from year to year, the number of tractor passes for the RT plots was always reduced by 40% or more relative to the ST plots (Mitchell et al., 2012).

The soil type at the study site is a Panoche clay loam (fine-loamy, mixed, superactive, thermic Typic Haplocambids) which is representative for much of California's Central Valley. For the first 12 years of the conservation agriculture experiment (between 2000 and 2012), tomato and cotton were grown in rotation, followed by a rotation of sorghum with garbanzo beans since 2012. All plots were irrigated by subsurface drip.

The cover crops were a mix of triticale (*Triticosecale Wittm*), cereal rye (*Secale cereale L.*), common vetch (*Vicia sativa*), radish (*Raphanus sativus*), and clover (*Trifolium incarnatum*) seeded in 20 cm rows at 89.2 kg ha- in late October. The cover crops are terminated in late March of the following year using a stalk chopper followed by disk incorporation in the ST system or sprayed with a 2% solution application of glyphosate after chopping and left on the surface as a mulch in the RT systems. Detailed description of the study site and management has been published in previous works (Mitchell, Shrestha, et al., 2016; Mitchell et al., 2015, 2017; Veenstra et al., 2006)

Figure 3-1 Location of the study site (red crosshair) in California's Central Valley (orange hatched fill) and some major cities of California (Topographic base map from the National Geographic Society (2013)).

3.2.2 Sampling

Sampling was done in mid-November 2017, several months after tillage in the ST treatment plots to avoid the immediate effects of tillage since we were primarily interested in the long-term effects of the treatments. Tillage operations have a transitory effect on porosity and associated soil hydraulic properties as the structures collapse, mainly driven by wetting

and drying cycles post tillage (Mapa et al., 1986; Or et al., 2000). The immediate alterations of tillage on soil porosity and hydraulic properties have been shown to diminish rapidly following by only a few wetting and drying cycles (Alletto et al., 2015; Green et al., 2003; Strudley et al., 2008).

Undisturbed soil samples from the top (0 – 5 cm) and subsurface (20 – 25 cm) layers were collected carefully using a 250 cm^3 volume sampling ring (8 cm diameter by 5 cm height). The depths were chosen to correspond with the depth disturbed by disking to incorporate residue in the ST plots (i.e., 0 – 20 cm depth) (Mitchell et al., 2015; Veenstra et al., 2006) and the deeper layer. Samples were collected along the strip ridges within the plots away from the trafficked furrows but slightly off-center to avoid drip irrigation tubes which were buried at the center of ridges. A total of 32 samples were collected by taking one surface, and one subsurface sample from four of the eight treatment replicate plots. This resulted in four replicates of surface and subsurface samples per treatment. The samples were stored at 4 °C before laboratory analysis.

3.2.3 Laboratory Measurements

The laboratory procedure followed for each soil core is illustrated in Figure 3-2. The saturated hydraulic conductivity (K_s) was measured using the falling-head method. For this method, soils were saturated by immersing sample cores in degassed, 0.01 M $CaCl_2$ solution so that the water level was close to the rim. K_s of the saturated soil was then measured by the falling-head method using the KSAT instrument (METER Group, Inc., Munich, Germany) by allowing 5 cm column of degassed, 0.01 M $CaCl_2$ solution to flow through the soil core. The set up was such so that the flow direction was downward. Following the K_s measurement, soil WRC and HCF data were determined simultaneously using the evaporation method as developed for the HYPROP instrument (METER Group, Inc., Munich, Germany). The HYPROP simultaneously measures, at high frequency (10 min), suction inside the soil cores at two different depths along with weight loss while saturated soil cores dry. This allows for the calculation of WRC, $\theta(\psi)$, and HCF, $K(\psi)$. Following the HYPROP measurements, soil water retention in the range from 10^3 to 10^6 cm was determined by using the WP4C instrument (Decagon Devices, Inc, Pullman, WA, USA).

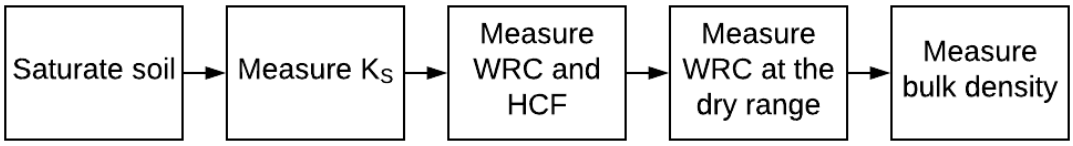

Figure 3-2 Flowchart of the measurement steps for the soil core samples.

We use the conventional definition for field capacity (θ_{FC}) and permanent wilting point (θ_{PWP}) as the volumetric water content with the corresponding volume of water retained in the soil at −33 kPa and −1,500 kPa suction, respectively. θ_{FC} and θ_{PWP} are approximations of water retained after internal drainage has ceased, and the soil water content limit at which plants cannot recover from turgidity, respectively (Hillel, 1998). We calculated plant available water (PAW) as the difference between θ_{FC} and θ_{PWP}. In addition to the saturated hydraulic conductivity, we also calculated the unsaturated hydraulic conductivity near field capacity water content at -10 kPa.

Throughout this manuscript, the term water suction, h, is used to represent the soil water matric potential, ψ, such that $h = -\psi$ (cm).

3.2.4 Soil Porosity Determination

Soil bulk density (ρ_b) was determined using the standard core method (Grossman & Reinsch, 2002). Total soil porosity (ϕ) was calculated from bulk density as Equation 4.1.

$$\phi = 1 - \frac{\rho_b}{\rho_p} \tag{4.1}$$

where ρ_p is the particle density of soil, taken as 2.65 g cm^{-3}.

The effective pore size distribution (PSD) was estimated from the slope of the WRC using the differential water capacity (Klute, 1986). For this, we followed procedures similar to Pires et al. (2017). The WRC, or $\theta(h)$ was first transformed into $S(r)$ by converting θ into effective saturation (S) using Equation 4.2 and approximating pore radius (r) using Equation 4.3).

For pre-determined values of θ, S was calculated as:

$$S = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{4.2}$$

where θ_s and θ_r are the saturated and residual volumetric water contents estimated from a bimodal constrained van Genuchten model fit (Durner, 1994) of measured WRC.

The corresponding draining pore radius, r, was approximated from h by the capillary pressure function:

$$r = \frac{2\gamma \cos(\beta)}{\rho_w g h} = \frac{1490}{h} \tag{4.3}$$

where r [µm] is pore radius, h [cm] is the suction, γ is the surface tension between water and air (72.9 dyn cm^{-1}), $\beta$is the contact angle (assumed 0), ρ_w is the density of water (1 g cm^{-3}), and g is the acceleration due to gravity (980 cm s^{-2}).

Prior to calculating PSD, we fitted the $S(r)$ with a cubic smoothing spline to remove noisy observations and to smooth data (Kastanek & Nielsen, 2001). PSD curves were then calculated as defined by (Equation 4.4)

$$f_p(\ln r) = -\frac{dS}{d\ln r} \tag{4.4}$$

where f_p [-] is the probability density function of effective pore sizes.

3.2.5 Soil Water Storage Simulations

One-dimensional numerical modeling was done using HYDRUS 1D software to analyze soil water storage for the different treatments using the soil physical characteristics as determined above. The HYDRUS 1D software simulates soil water flow using the Richards equation and incorporates a sink term to account for water uptake by plant roots, as shown in Equation 4.5 (Simunek et al., 2013).

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x}\left[K\left(\frac{\partial h}{\partial x} + \cos\alpha\right)\right] - S_r \tag{4.5}$$

where t is time, x is the spatial coordinate, α is the angle between the flow direction and vertical axis (0° for vertical flow), and S_r is the sink term.

We simulated the fate of 4 cm irrigation applied at a rate of 1.3 cm h^{-1} in each of the 16 sampled plots over time. We calculated the loss and distribution of the water in the soil profile. The simulation was set up over a 100 cm deep soil made up of two materials with different soil hydraulic properties, the top 20 cm had the hydraulic properties from the top layer soils (0 – 5 cm) and the bottom 80 cm subsoil had the hydraulic properties of the subsurface layer soils (20 – 25 cm) sample. The measured soil hydraulic properties of WRC and HCF were supplied to the model via HYDRUS 1D's look-up table option. For water distribution accounting purposes, the soil profile was split into three layers: top (0 – 19 cm), middle (20 – 39 cm) and bottom (40 – 100 cm) layers. Figure 3-3 illustrates the soil profile setup for the simulation. Atmospheric boundary condition was set for the surface layer and a free drainage lower boundary for the bottom layer. For the computation of evaporation and transpiration, daily meteorological parameters of radiation, minimum and maximum air temperature, relative humidity and wind speed for a week during May 2018 were acquired from a weather station near the study site (CIMIS Five Points Station, https://cimis.water.ca.gov/). A tomato plant with Feddes root water uptake distribution linearly distributed from the soil surface to 30 cm rooting depth was used with the default parameters supplied by HYDRUS 1D software. The root distribution was then calculated from Equation 4.6.

$$b = 1 - \frac{z}{31} \tag{4.6}$$

where b is the normalized root distribution, and z is soil depth in cm.

The simulation was initialized by running repeated 2 cm weekly irrigation for 14 weeks prior to the final simulation of 4 cm irrigation application at a rate of 1.3 cm h^{-1}. Analysis of water capture and storage following the final irrigation application was analyzed by calculating the change in water storage following irrigation from Equation 4.7

$$\Delta W(t) = W(t) - W(t_0) \quad (4.7)$$

where W [cm] is water amount across the soil profile, t is simulation time where t_0 is the end of irrigation application.

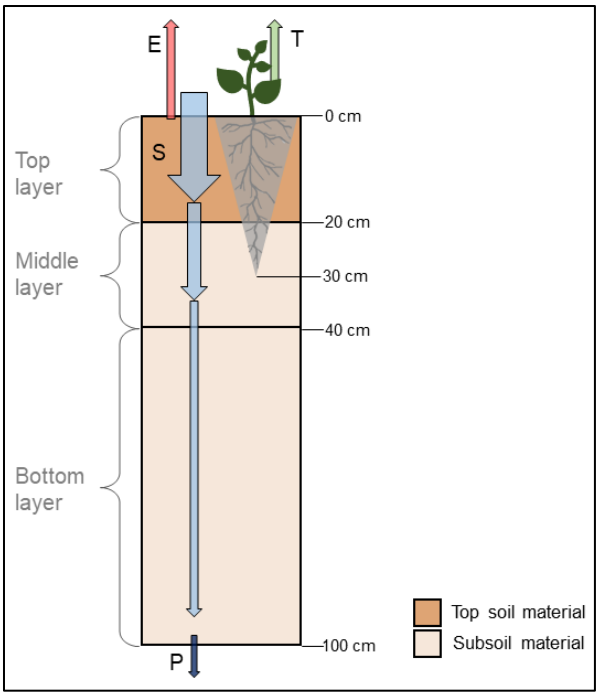

Figure 3-3 Schematic representation of the domain geometry and material distribution used for HYDRUS 1D simulation model. Arrows represent the water flow pathways; T, E, S, and P stand for transpiration, evaporation, soil storage, and percolation, respectively.

3.2.6 Statistical Analysis

All quantitative results are expressed as means of four replicates ± standard error unless otherwise indicated. Differences in means were tested by analysis of variance (ANOVA) and pairwise comparison of treatments done using Tukey's honest significant difference (HSD) test at $p < 0.15$ significance level unless otherwise stated. Hydraulic conductivity values were log-transformed before statistical analysis to make their distribution more normal. The normality of the data and the homogeneity of variances was checked using Shapiro–Wilk's and Levene's tests, respectively. All statistical analyses were performed using R statistical software (R Core Team, 2019).

3.3 Results and Discussion

Hydraulic properties of HCF and WRC were measured for all 32 samples by using KSAT, HYPROP, and WP4C instruments. Figure 3-4 shows the laboratory measurement results for one of the sampled soilsFigure A0-2. The complete measured HCF and WRC for all the soils is provided in Appendix B (Figures A0-4 to A0-6).

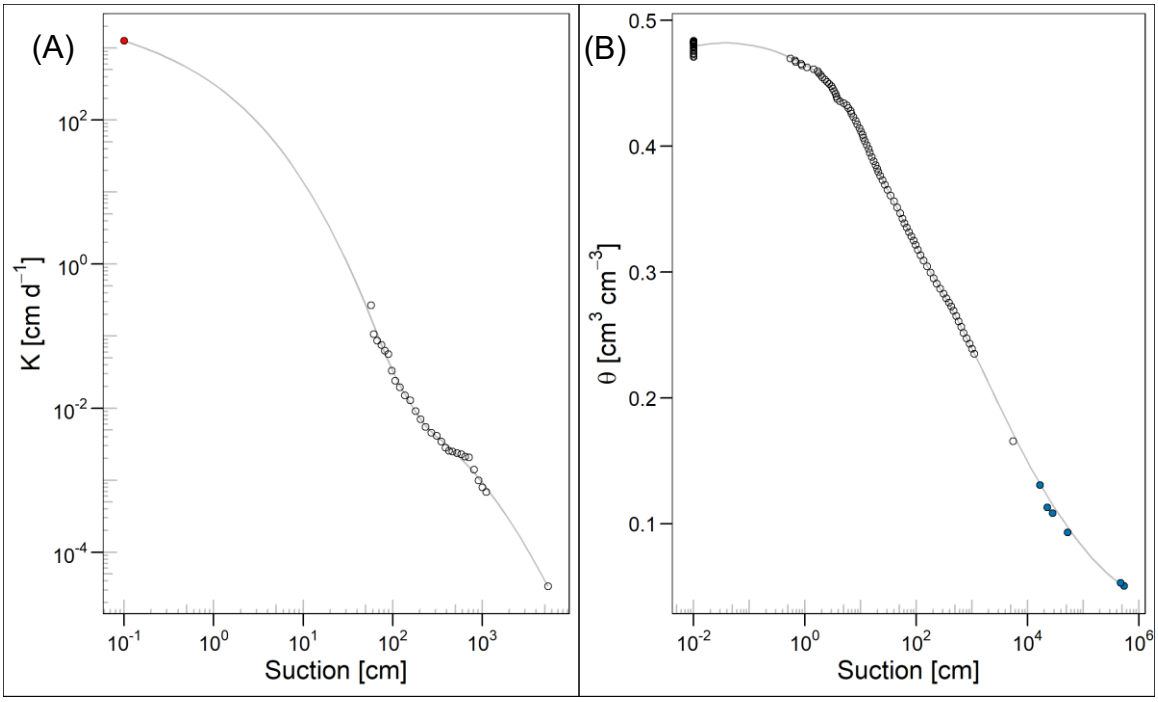

Figure 3-4 Measured (A) hydraulic conductivity function and (B) water retention curve for one of the topsoil RT-CC plots. The red-filled circle was measured using KSAT instrument, open circles were measured using HYPROP, and blue-filled circles were measured using WP4C instrument.

3.3.1 Soil Physical and Hydraulic Properties

The mean ρ_b across all treatments for the top and subsurface layer soils was 1.19 and 1.46 g cm^{-3}, respectively. Among the treatments, there was no statistically significant difference ($p < 0.15$) in ρ_b and ϕ (Figure 3-5). One of the concerns of RT practice is that it may lead to soil consolidation and increase in compaction because of the lack of intensive tillage (Blanco-Canqui & Ruis, 2018; Moret & Arrúe, 2007). Compaction reduces soil pore volume and affects soil fertility by reducing water flow and aeration, which negatively affect soil biological activity and redox potential (Vereecken et al., 2016). Our findings suggest that continued long-term RT did not lead to increased compaction. In contrast, changes in PSD (see section 3.3.2) showed that the RT treatments increased the PSD in a manner that suggested a better-developed soil structure with primary and secondary structures.

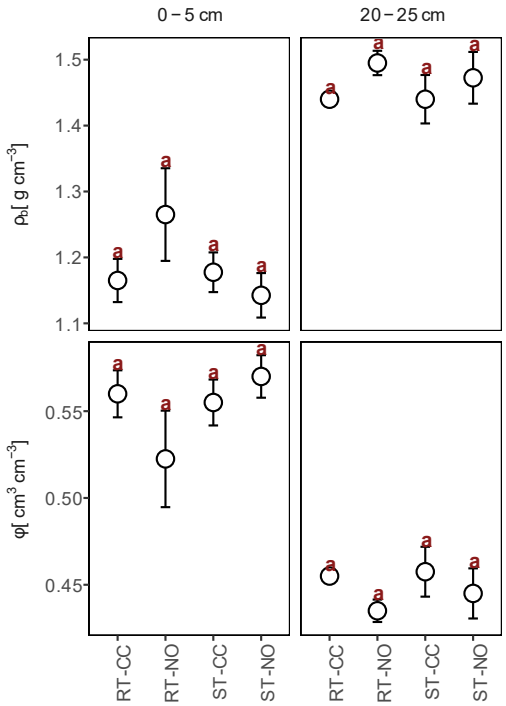

Figure 3-5 Treatment means of bulk density and total porosity. Bars indicate standard errors.

Figure 3-6 shows the values of some important soil hydraulic variable for the different treatments.

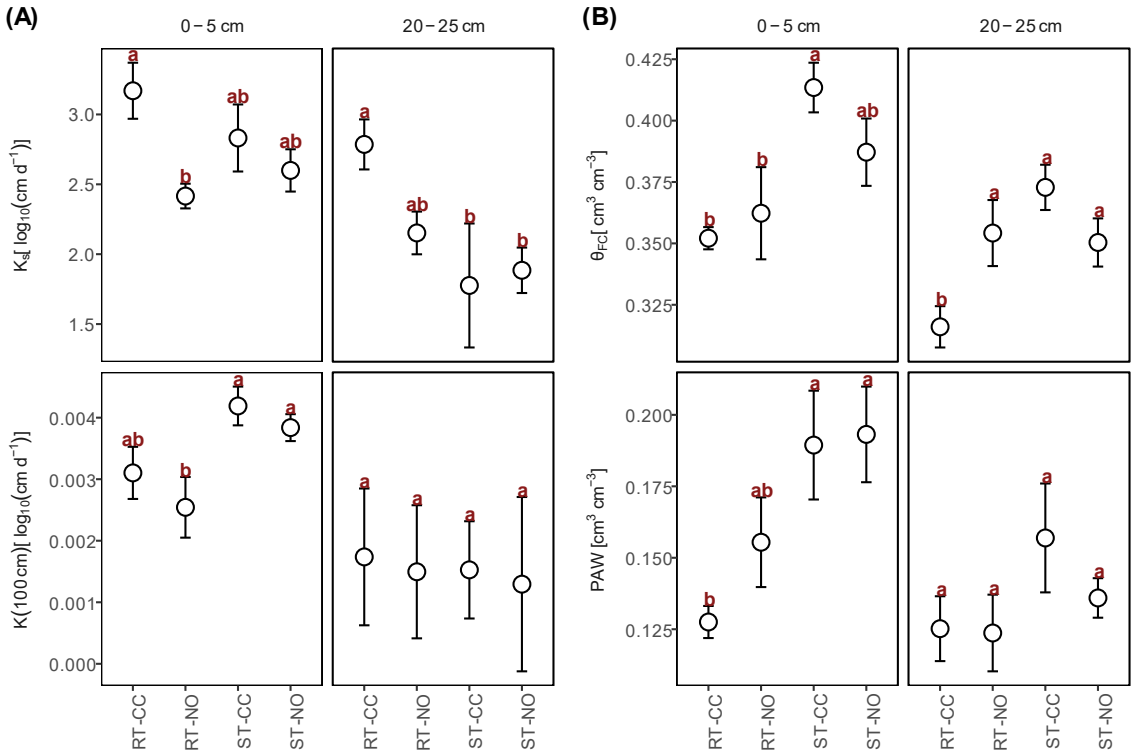

Figure 3-6 Treatment means of some characteristic soil (A) conductivity and (B) retention variables. Bars indicate standard errors. Different letters indicate statistically significant different means at $p < 0.15$ tested by Tukey's HSD test.

The RT treatments had lower θ_{FC} compared to ST treatments. The larger value of θ_{FC} for ST plots are consistent with a more loose soil due to tillage increasing the capillary size pores. The θ_{FC} for the top layer RT soils were lower by more than 5 % volumetric water content ($p<0.07$) compared to ST-CC. The ST-NO treatments had intermediate values which were not significantly different ($p<0.15$) from all other treatments. The θ_{FC} showed similar trends for the subsurface layer soils but with smaller magnitudes of differences. CC appeared to enhance the effects of RT in terms of θ_{FC} and PAW of topsoil layers. The RT-NO top layer soils showed values between RT-CC and the ST soils. The top layers of RT-CC plots showed a statistically significant decrease in PAW ($p < 0.1$) compared to the ST treatments. This was mainly driven by the lower θ_{FC} values of the RT treatments. On both layers, the CC treatment increased θ_{FC} on of ST soils but had the opposite effect on the RT soils. While some studies reported an increase in θ_{FC} and PAW with CC (A. D. Basche, Kaspar, et al., 2016; Bilek, 2007; Villamil et al., 2006), our findings are consistent with the observations from a recent meta-analysis of 93 paired observations of CC (A. Basche & DeLonge, 2017) which showed that CC had no effect on ϕ for treatments practiced longer than 7 years or clay contents > 25 % which match the parameters of our study site. Our findings also agree with the findings of Basche and DeLonge (2017) in terms of θ_{FC}, they find that while long-term CC tends to increase θ_{FC}, it actually tends to decrease it for soils with >25% clay. Our results showed that while this was the case with ST, it was not the

case for RT. For the subsurface layer of RT treatments, θ_{FC} was significantly lower ($p< 0.1$) the RT-CC compared with RT-NO treatments.

CC appears to have more effect on K_s than the tillage treatment for the top layer soils. This is consistent with the increase in infiltration reported previously for our soils by Mitchell et al. (2017). They found that CC increased infiltration by 2.8 times. They suggest several possible explanations for this including increased slaking associated with ST, better formation of macropores, and better continuity of soil pores possibly due to better-established soil structure and biology (Pires et al., 2017; Schwen et al., 2011). The RT-CC plots showed statistically significant ($p < 0.1$) higher K_s than RT-NO. The ST plots showed K_s midway between the RT-NO and RT-CC treatments. The fact that K_s of RT-NO treatments is lower even more than ST plots suggests that CC is even more important when RT is practiced to maintain larger transmission pores without tillage. The effect of CC on ST treatments was small and not statistically significant ($p < 0.15$). K at 100 cm suction, $K(100\ cm)$, is controlled by smaller pores as opposed to K_s. The RT treatments had lower $K(100\ cm)$ compared to ST plots, which implies that on unsaturated soils, the RT treatments will lose water more slowly than ST plots. This could possibly mitigate the impact of reduced θ_{FC} in the RT treatments and lead to an increase of water availability to plants. Results of the numerical simulation also seem to support this (see section 3.3.3).

3.3.2 Pore Size Distribution

The PSD for the treatments are shown in Figure 3-7 and more detailed PSD curves showing individual replicate curves are provided in Appendix B, Figure A0-6. Greenland (1977) suggests soil pore size classification based on equivalent cylindrical diameter into three groups as transmission (50 – 500 μm), storage (0.5 – 50 μm) and residual pores (< 0.5 μm). Larger transmission pores are important for infiltration, drainage, and aeration while smaller storage pores are important in retaining water. Increased aeration of soil is beneficial for many soil processes including healthy soil organic matter cycling (Janzen, 2015; Lehmann & Kleber, 2015) and other biogeochemical processes (Ekschmitt et al., 2008; Schmidt et al., 2011).

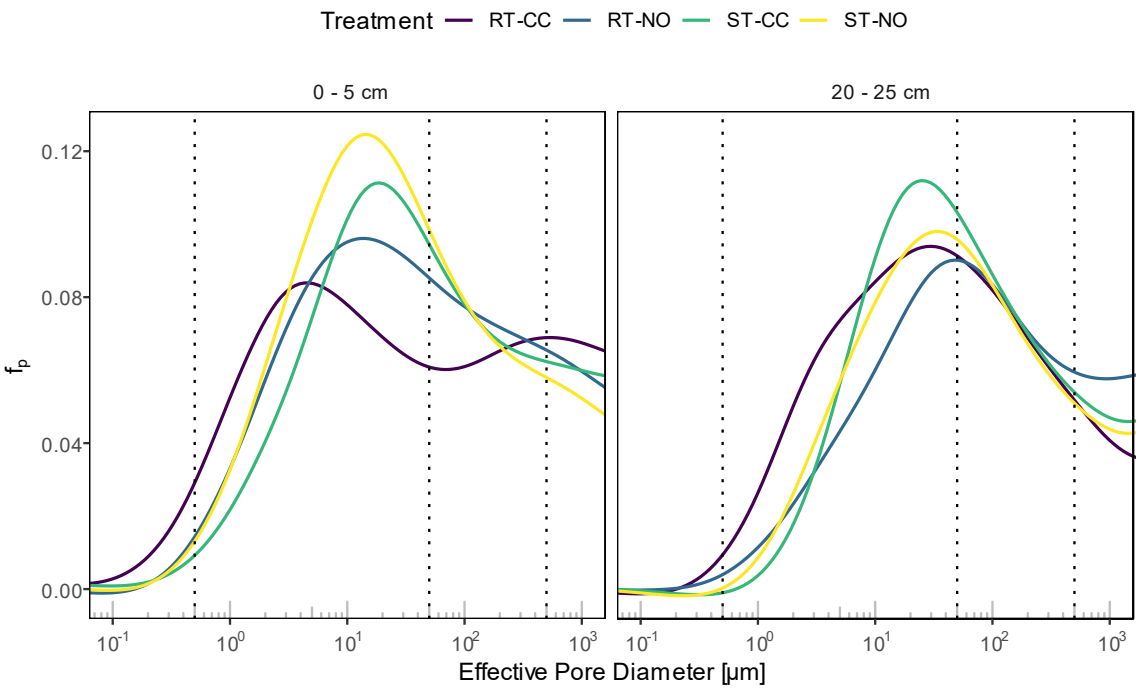

Figure 3-7 Treatment means of the effective pore size distribution for the top (0 – 5 cm) and subsurface (20 – 25 cm) layers. Vertical dotted lines indicate pore diameter sizes of 0.5, 50, and 500 µm.

The width of PSD is indicative of soil with a more developed structure with a heterogeneous mix of pore sizes. The width of PSD for the top layer was in the order of RT-CC > RT-NO > ST-CC > ST-NO. The PSD shows that topsoil RT-CC has a wider PSD with more pores of larger diameter and a notable bimodal distribution which was not present in the other treatments. Several studies have found an increase in the proportion of larger pores in RT treatments (Pires et al., 2017; Tavares Filho & Tessier, 2009). The PSD curves for RT-CC treatment also show an increase in micropores compared to the other treatments.

The maximum PSD value for the top soils occurred between 10 and 20 µm diameter pores except for RT-CC soils which occurred around 4 and 500 µm. The reason for the abundance of small and large pores suggests the formation of tightly packed aggregates with smaller pores and larger interaggregate pores between them. For the subsurface soil, the combined effect of RT and CC increased the PSD while RT without CC treatments showed a much narrower PSD, narrower than ST treatments. Plant roots are important actors in soil structure development, they enhance aggregation by compacting soils through growth and exudation of segmenting materials, and also fragmenting aggregates to create larger interaggregate pores (Angers & Caron, 1998; Jarvis, 2007). Given the reduced tillage in the RT treatments, it could be that CC play a more critical role in forming more diverse aggregate sizes and wider PSD. The effect of the CC species should also be considered in this interpretation since it has recently been showed that the effect of CC on soil structure and porosity varies significantly with root morphology and architecture of the CC plant (Bacq-Labreuil et al., 2019).

3.3.3 Soil Storage Simulations

Figure 3-8 shows the soil moisture over time. The simulation was initialized by running repeated 2 cm weekly irrigation for 14 weeks prior to the final 4 cm irrigation application at a rate of 1.3 cm h^{-1}. The RT-CC consistently retained the most water in the top 20 cm layer and ST-NO the least. The RT-NO and ST-CC soils had intermediate storage.

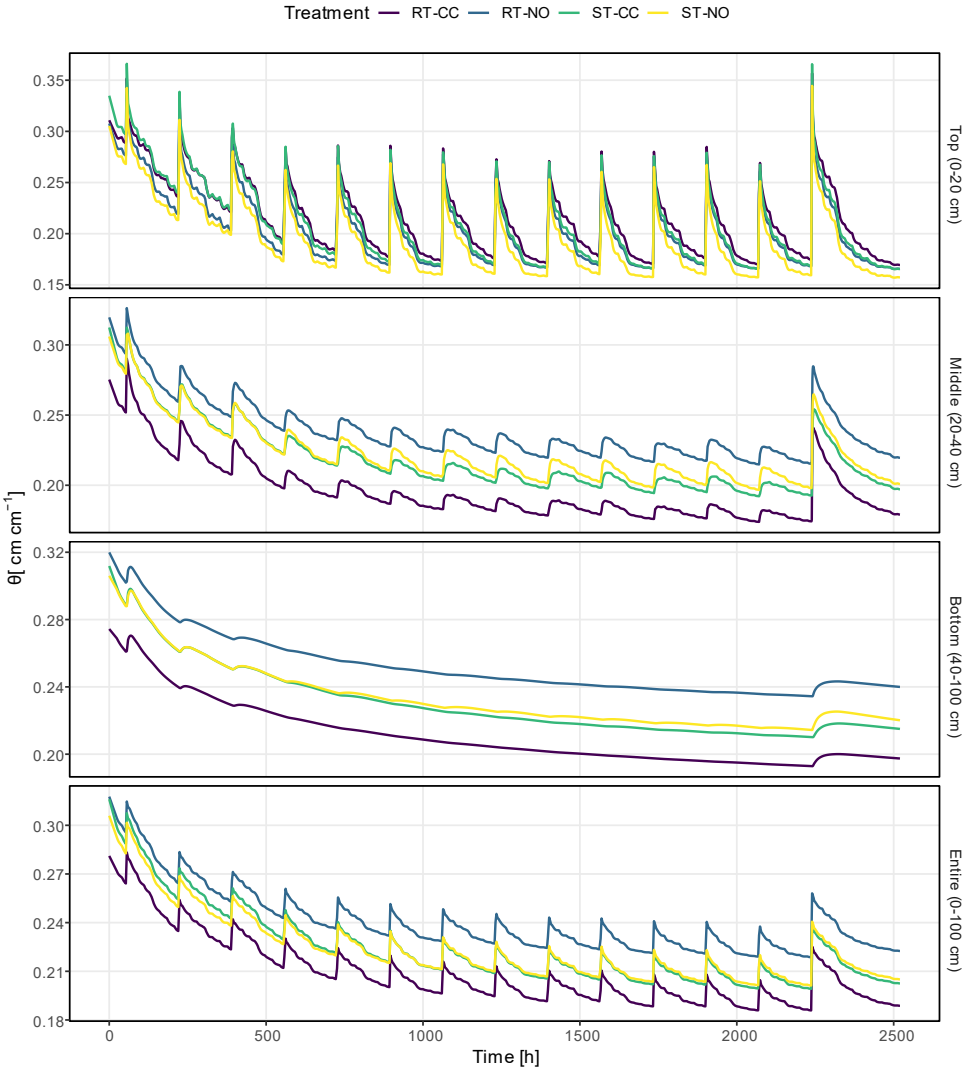

Figure 3-8 Water content changes of the three layers and the entire profile over simulation time. The simulation was initialized by repeated 2 cm weekly irrigation for 14 weeks before applying the final 4 cm irrigation.

The change in water storage following 4 cm water irrigation is shown in Figure 3-9. The top layer starts to lose water instantly by evapotranspiration and drainage to the lower layers. The water storage of the top layer was in the order of RT-CC > RT-NO > ST-NO >ST-CC over the weeks following irrigation. The distribution of water in the deeper profiles shows that the top and middle layers of ST-NO soils drained the most. For the ST-CC soils, the top layer drained water in a similar pattern to the ST-NO soils. However, the

middle layer retained most of the water. Depending on whether plant roots can reach this layer, this could be a positive since the water retained in the middle layer is less affected by evaporative loss than the top layer. The water storage changes of middle and bottom layers for the RT treatments were similar. Plots of the actual water storage (depth of water) for five days following irrigation for the three layers and the entire profile shown in Appendix B, Figure A0-7.

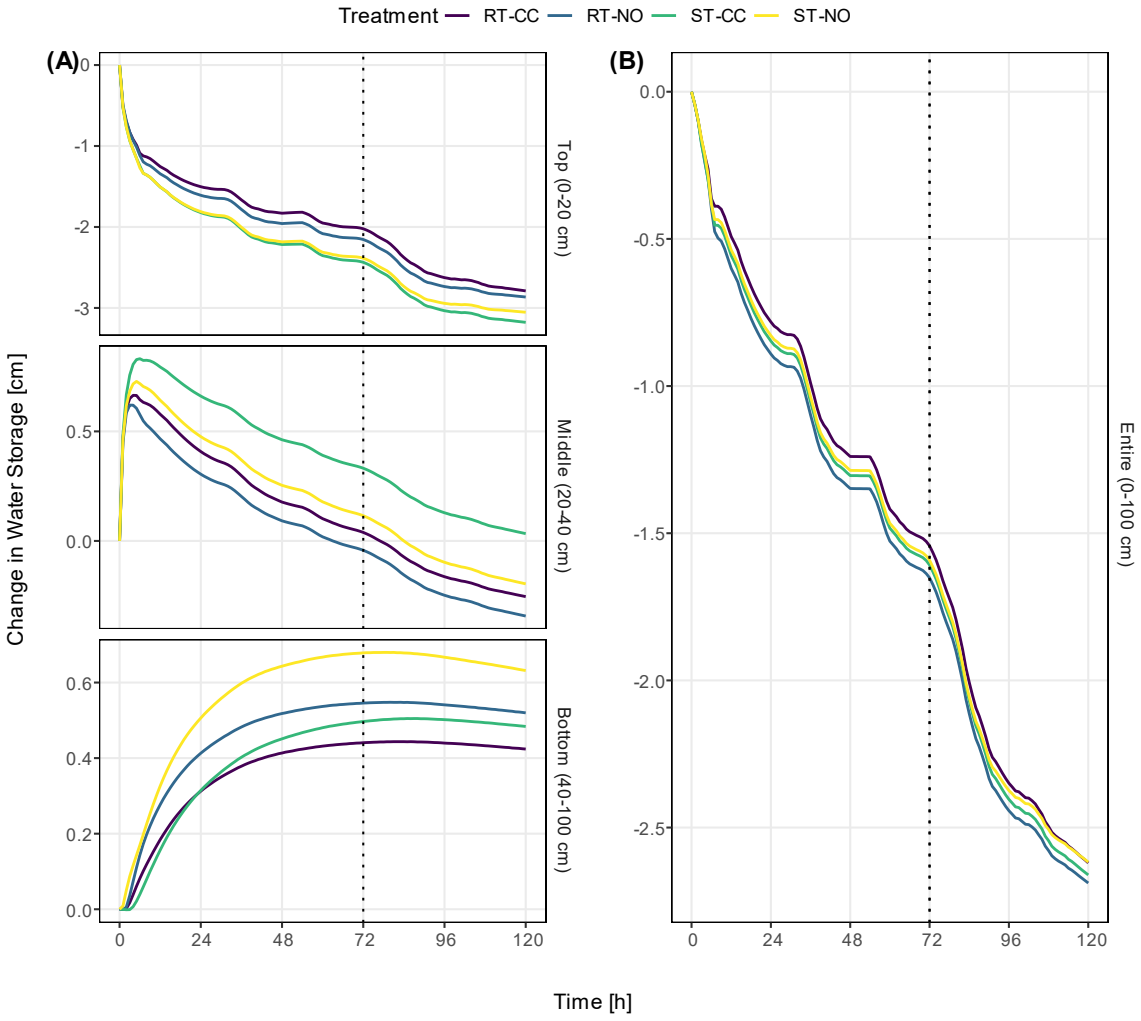

Figure 3-9 Change in water storage following the 4 cm irrigation (A) in the three layers and (B) in the entire profile. The dotted vertical line indicates the 3rd day after irrigation.

The dynamic field capacity is defined as the "*amount of water held in soil after excess gravitational water has drained away and after the rate of downward movement has materially decreased*" (Veihmeyer & Hendrickson, 1931) and is traditionally taken as the water content after three (or sometimes even five) days (Assouline & Or, 2014; Twarakavi et al., 2009). In our simulation, the rate of water drainage for the top and middle layers had

significantly decreased after three days. Furthermore, there was no change in the relative amount of water storage among the treatments for the following six days (Figure 3-9).

To approximate this and to get an absolute amount of water volume retained in the top layer, we compared the average θ for each layer and the amount of water retained from the 4 cm irrigation water for the top layers three days after irrigation Figure 3-10. Both the RT treatments had the highest water storage. This finding is not consistent with the θ_{FC} and the PAW estimated from the traditional steady-state measures (compare with section 3.3.1). The dynamic water storage is the interplay between WRC and HCF and possibly captures soil hydrology more accurately.

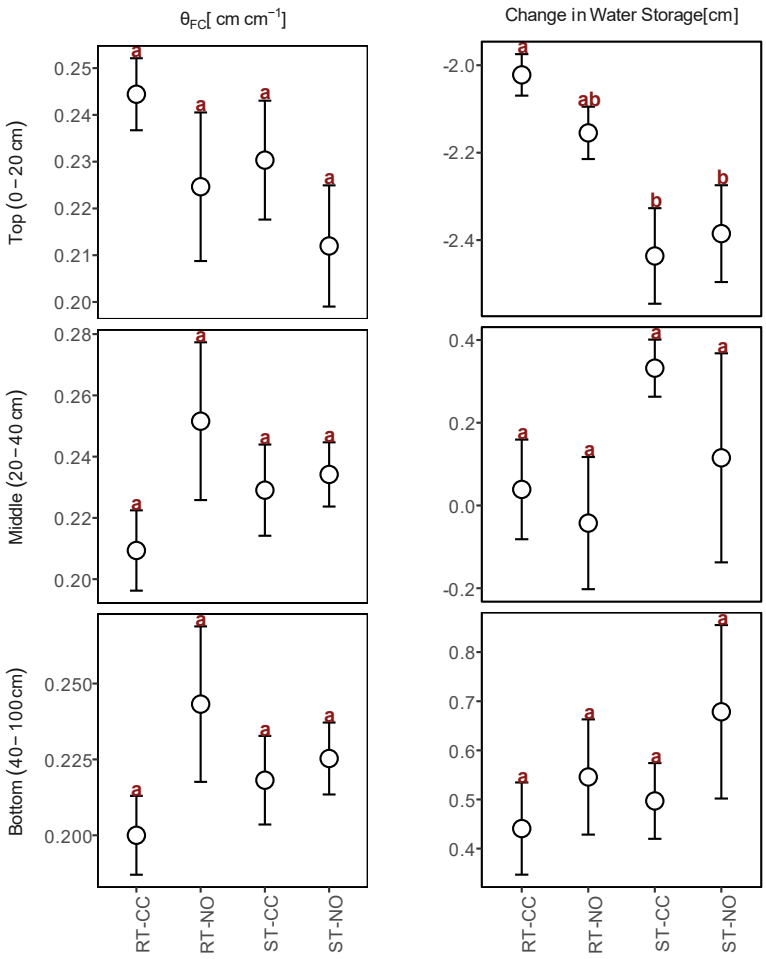

Figure 3-10 Dynamic field capacity (θ_{FC}) and amount of water storage change for the three layers, three days after irrigation. Bars indicate standard errors. Different letters indicate statistically significant different means at $p < 0.15$ tested by Tukey's HSD test.

3.4 Conclusion

The long-term reduced tillage (RT) and cover crop (CC) practices had an impact on soil pore size distribution (PSD). The RT and CC practiced independently led to a mild increase in PSD range and had small or no effect on the measured soil hydraulic properties and simulated water dynamics. On the plots where RT and CC were practiced the changes in soil structure and hydraulic properties were most pronounced. RT with CC led to development of bimodal pore size distribution in the top (0 –5 cm) soils with the modes of the PSD around 4 and 500 μm diameter sizes which are in the storage and transmission pore sizes. While ST is done to improve soil structure for crops and overcome the compaction of the topsoil layer, its effect is transitory. Our results suggest that in the longer-term, RT and CC increase soil aggregation and the proportion of larger pores while also maintaining total porosity.

CC appeared to increase the saturated hydraulic conductivity, particularly in the RT plots. For the top layer soils (0 – 5 cm), the saturated hydraulic conductivity in the RT-CC treatments was significantly higher ($p < 0.15$) than in the RT only treatments and for the subsoil layer (20 – 25 cm), significantly higher ($p < 0.15$) than the ST treatments.

The measured water retention suggested a decrease in soils ability to store water. The RT with CC decreased the calculated plant available water (PAW) and water content at field capacity ($\theta(300\ cm)$). While the static measures of field capacity and PAW indicate soil's ability to store water, the actual water storage in soils is the result of the interaction between soil's water retention characteristics and its hydraulic conductivities. Both the water retention and conductivity were accounted for in the HYDRUS irrigation simulation. The results showed that when both retention and conductivity are considered together, the RT plots with CC have significantly increased ability to store water compared to ST plots ($p < 0.15$) while the RT plots without CC have intermediate storage capacity.

The changes in PSD and water storage associated with RT and CC are likely to have a beneficial effect on soil health by providing a more abundant supply of air and water to deeper soil and retaining more soil moisture.

Acknowledgments

I want to thank Professors Jeffrey P. Mitchell and Jan Hopmans (UC Davis) for their guidance and insightful comments. I am thankful for the Conservation Agriculture Systems Innovation (CASI) and the California Department of Water Resources for making this work possible.

References

Abdollahi, L., Schjønning, P., Elmholt, S., & Munkholm, L. J. (2014). The effects of organic matter application and intensive tillage and traffic on soil structure formation and stability. *Soil and Tillage Research*, *136*, 28–37. https://doi.org/10.1016/j.still.2013.09.011

Alletto, L., Pot, V., Giuliano, S., Costes, M., Perdrieux, F., & Justes, E. (2015). Temporal variation in soil physical properties improves the water dynamics modeling in a conventionally-tilled soil. *Geoderma*, *243–244*, 18–28. https://doi.org/10.1016/j.geoderma.2014.12.006

Alvarez, R., & Steinbach, H. S. (2009). A review of the effects of tillage systems on some soil physical properties, water content, nitrate availability and crops yield in the Argentine Pampas. *Soil and Tillage Research*, *104*(1), 1–15. https://doi.org/10.1016/J.STILL.2009.02.005

Angers, D. A., & Caron, J. (1998). Plant-induced Changes in Soil Structure: Processes and Feedbacks. *Biogeochemistry*, *42*(1/2), 55–72. https://doi.org/10.1023/A:1005944025343

Ashworth, A. J., DeBruyn, J. M., Allen, F. L., Radosevich, M., & Owens, P. R. (2017). Microbial community structure is affected by cropping sequences and poultry litter under long-term no-tillage. *Soil Biology and Biochemistry*, *114*, 210–219. https://doi.org/10.1016/J.SOILBIO.2017.07.019

Assouline, S., & Or, D. (2014). The concept of field capacity revisited: Defining intrinsic static and dynamic criteria for soil internal drainage dynamics. *Water Resources Research*, *50*(6), 4787–4802. https://doi.org/10.1002/2014WR015475

Bacq-Labreuil, A., Crawford, J., Mooney, S. J., Neal, A. L., & Ritz, K. (2019). Cover crop species have contrasting influence upon soil structural genesis and microbial community phenotype. *Scientific Reports*, *9*(1), 7473. https://doi.org/10.1038/s41598-019-43937-6

Baker, J. B., Southard, R. J., & Mitchell, J. P. (2005). Agricultural Dust Production in Standard and Conservation Tillage Systems in the San Joaquin Valley. *Journal of Environment Quality*, *34*(4), 1260. https://doi.org/10.2134/jeq2003.0348

Basche, A. D., Archontoulis, S. V., Kaspar, T. C., Jaynes, D. B., Parkin, T. B., & Miguez, F. E. (2016). Simulating long-term impacts of cover crops and climate change on crop production and environmental outcomes in the Midwestern United States. *Agriculture, Ecosystems & Environment*, *218*, 95–106. https://doi.org/10.1016/J.AGEE.2015.11.011

Basche, A. D., Kaspar, T. C., Archontoulis, S. V., Jaynes, D. B., Sauer, T. J., Parkin, T. B., & Miguez, F. E. (2016). Soil water improvements with the long-term use of a winter rye cover crop. *Agricultural Water Management*, *172*, 40–50.

https://doi.org/10.1016/j.agwat.2016.04.006

Basche, A., & DeLonge, M. (2017). The Impact of Continuous Living Cover on Soil Hydrologic Properties: A Meta-Analysis. *Soil Science Society of America Journal*, *81*(5), 1179. https://doi.org/10.2136/sssaj2017.03.0077

Bilek, M. (2007). *Winter annual rye cover crops in no-till grain crop rotations: impacts on soil physical properties and organic matter*. University of Maryland, College Park. Retrieved from https://drum.lib.umd.edu/handle/1903/7268

Blanco-Canqui, H., & Ruis, S. J. (2018). No-tillage and soil physical environment. *Geoderma*, *326*, 164–200. https://doi.org/10.1016/j.geoderma.2018.03.011

Burr-Hersey, J. E., Mooney, S. J., Bengough, A. G., Mairhofer, S., & Ritz, K. (2017). Developmental morphology of cover crop species exhibit contrasting behaviour to changes in soil bulk density, revealed by X-ray computed tomography. *PLOS ONE*, *12*(7), e0181872. https://doi.org/10.1371/journal.pone.0181872

Duchene, O., Vian, J.-F., & Celette, F. (2017). Intercropping with legume for agroecological cropping systems: Complementarity and facilitation processes and the importance of soil microorganisms. A review. *Agriculture, Ecosystems & Environment*, *240*, 148–161. https://doi.org/10.1016/J.AGEE.2017.02.019

Durner, W. (1994). Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resources Research*, *30*(2), 211–223. https://doi.org/10.1029/93WR02676

Ekschmitt, K., Kandeler, E., Poll, C., Brune, A., Buscot, F., Friedrich, M., … Wolters, V. (2008). Soil-carbon preservation through habitat constraints and biological limitations on decomposer activity. *Journal of Plant Nutrition and Soil Science*, *171*(1), 27–35. https://doi.org/10.1002/jpln.200700051

Fernandez, A. L., Sheaffer, C. C., Wyse, D. L., Staley, C., Gould, T. J., & Sadowsky, M. J. (2016). Structure of bacterial communities in soil following cover crop and organic fertilizer incorporation. *Applied Microbiology and Biotechnology*, *100*(21), 9331–9341. https://doi.org/10.1007/s00253-016-7736-9

Finney, D. M., Buyer, J. S., & Kaye, J. P. (2017). Living cover crops have immediate impacts on soil microbial community structure and function. *Journal of Soil and Water Conservation*, *72*(4), 361–373. https://doi.org/10.2489/jswc.72.4.361

González-Sánchez, E. J., Kassam, A., Basch, G., Streit, B., Holgado-Cabrera, A., & Triviño-Tarradas, P. (2016). Conservation Agriculture and its contribution to the achievement of agri-environmental and economic challenges in Europe. *AIMS Agriculture and Food*. https://doi.org/10.3934/agrfood.2016.4.387

Green, T. R., Ahuja, L. R., & Benjamin, J. G. (2003). Advances and challenges in predicting agricultural management effects on soil hydraulic properties. *Geoderma*,

116(1–2), 3–27. https://doi.org/10.1016/S0016-7061(03)00091-0

Greenland, D. J. (1977). Soil Damage by Intensive Arable Cultivation: Temporary or Permanent? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *281*(980), 193–208. https://doi.org/10.1098/rstb.1977.0133

Grossman, R. B., & Reinsch, T. G. (2002). Bulk Density and Linear Extensibility. In J. H. Dane & G. C. Topp (Eds.), *Methods of Soil Analysis, Part 4--Physical Methods* (pp. 201–228). Madison, Wisconsin: Soil Science Society of America. https://doi.org/10.2136/sssabookser5.4.c9

Hillel, D. (1998). *Environmental Soil Physics*. San Diego, CA: Academic Press.

Hudson, B. D. (1994). Soil organic matter and available water capacity. *Journal of Soil and Water Conservation*, *49*(2), 189–194. Retrieved from http://www.jswconline.org/content/49/2/189.abstract

Janzen, H. H. (2015). Beyond carbon sequestration: soil as conduit of solar energy. *European Journal of Soil Science*, *66*(1), 19–32. https://doi.org/10.1111/ejss.12194

Jarvis, N. J. (2007). A review of non-equilibrium water flow and solute transport in soil macropores: principles, controlling factors and consequences for water quality. *European Journal of Soil Science*, *58*(3), 523–546. https://doi.org/10.1111/j.1365-2389.2007.00915.x

Johnson, A. M. M., & Hoyt, G. D. D. (1999). Changes to the soil environment under conservation tillage. *HortTechnology*, *9*(3), 380–393. https://doi.org/10.21273/HORTTECH.9.3.380

Kassam, A., Friedrich, T., & Derpsch, R. (2019). Global spread of Conservation Agriculture. *International Journal of Environmental Studies*, *76*(1), 29–51. https://doi.org/10.1080/00207233.2018.1494927

Kastanek, F. J., & Nielsen, D. R. (2001). Description of Soil Water Characteristics Using Cubic Spline Interpolation. *Soil Science Society of America Journal*, *65*(2), 279. https://doi.org/10.2136/sssaj2001.652279x

Lal, R., Reicosky, D. C., & Hanson, J. D. (2007). Evolution of the plow over 10,000 years and the rationale for no-till farming. *Soil and Tillage Research*, *93*(1), 1–12. https://doi.org/10.1016/j.still.2006.11.004

Lehmann, J., & Kleber, M. (2015). The contentious nature of soil organic matter. *Nature*, *528*(7580), 60–68. https://doi.org/10.1038/nature16069

Li, Y., Chang, S. X., Tian, L., & Zhang, Q. (2018). Conservation agriculture practices increase soil microbial biomass carbon and nitrogen in agricultural soils: A global meta-analysis. *Soil Biology and Biochemistry*, *121*, 50–58. https://doi.org/10.1016/j.soilbio.2018.02.024

Liu, J., Macrae, M. L., Elliott, J. A., Baulch, H. M., Wilson, H. F., & Kleinman, P. J. A. (2019). Impacts of Cover Crops and Crop Residues on Phosphorus Losses in Cold Climates: A Review. *Journal of Environment Quality*, *0*(0), 0. https://doi.org/10.2134/jeq2019.03.0119

Madden, N. M., Southard, R. J., & Mitchell, J. P. (2008). Conservation tillage reduces PM10 emissions in dairy forage rotations. *Atmospheric Environment*, *42*(16), 3795–3808. https://doi.org/10.1016/J.ATMOSENV.2007.12.058

Mapa, R. B., Green, R. E., & Santo, L. (1986). Temporal Variability of Soil Hydraulic Properties with Wetting and Drying Subsequent to Tillage. *Soil Science Society of America Journal*, *50*(5), 1133. https://doi.org/10.2136/sssaj1986.03615995005000050008x

Martens, D. A. (2004). Nitrogen cycling under different soil management systems (pp. 143–192). https://doi.org/10.1016/s0065-2113(01)70005-3

Mitchell, J. P. (2016). Conservation agriculture: systems thinking for sustainable farming. *California Agriculture*, *70*(2), 53–55.

Mitchell, J. P., Carter, L. M., Reicosky, D. C., Shrestha, A., Pettygrove, G. S., Klonsky, K. M., … Dunning, L. (2016). A history of tillage in California's Central Valley. *Soil and Tillage Research*, *157*, 52–64. https://doi.org/10.1016/j.still.2015.10.015

Mitchell, J. P., Klonsky, K. M., Miyao, E. M., Aegerter, B. J., Shrestha, A., Munk, D. S., … Turini, T. A. (2012). Evolution of Conservation Tillage Systems for Processing Tomato in California's Central Valley. *HortTechnology*, *22*(5), 617–626.

Mitchell, J. P., Klonsky, K. M., Miyao, E. M., & Hembree, K. J. (2009). Conservation tillage tomato production in California's San Joaquin Valley. *Agriculture and Natural Resources*, (January). Retrieved from http://anrcatalog.ucanr.edu/Details.aspx?itemNo=8330

Mitchell, J. P., Pettygrove, G. S., Upadhyaya, S., Shrestha, A., Fry, R., Roy, R., … Hembree, K. (2009). *Classification of Conservation Tillage Practices in California Irrigated Row Crop Systems*. *Classification of Conservation Tillage Practices in California Irrigated Row Crop Systems*. University of California, Agriculture and Natural Resources. https://doi.org/10.3733/ucanr.8364

Mitchell, J. P., Reicosky, D. C., Kueneman, E. A., Fisher, J., & Beck, D. (2019). Conservation agriculture systems. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, *14*(001). https://doi.org/10.1079/PAVSNNR201914001

Mitchell, J. P., Shrestha, A., Dahlberg, J. A., Munk, D. S., & Hembree, K. J. (2016). Prospect of No-till Planting of Sorghum with and without Cover Cropping in the San Joaquin Valley. *Crop, Forage and Turfgrass Management*, *2*(1), 0. https://doi.org/10.2134/cftm2015.0208

Mitchell, J. P., Shrestha, A., Horwath, W. R., Southard, R. J., Madden, N. M., Veenstra, J. J., & Munk, D. S. (2015). Tillage and Cover Cropping Affect Crop Yields and Soil Carbon in the San Joaquin Valley, California. *Agronomy Journal*, *107*(2), 588. https://doi.org/10.2134/agronj14.0415

Mitchell, J. P., Shrestha, A., Mathesius, K., Scow, K. M., Southard, R. J., Haney, R. L., … Horwath, W. R. (2017). Cover cropping and no-tillage improve soil health in an arid irrigated cropping system in California's San Joaquin Valley, USA. *Soil and Tillage Research*, *165*, 325–335. https://doi.org/10.1016/j.still.2016.09.001

Moret, D., & Arrúe, J. L. (2007). Dynamics of soil hydraulic properties during fallow as affected by tillage. *Soil and Tillage Research*, *96*(1–2), 103–113. https://doi.org/10.1016/j.still.2007.04.003

Naab, J. B., Mahama, G. Y., Yahaya, I., & Prasad, P. V. V. (2017). Conservation Agriculture Improves Soil Quality, Crop Yield, and Incomes of Smallholder Farmers in North Western Ghana. *Frontiers in Plant Science*, *8*, 996. https://doi.org/10.3389/fpls.2017.00996

Or, D., Leij, F. J., Snyder, V., & Ghezzehei, T. A. (2000). Stochastic model for posttillage soil pore space evolution. *Water Resources Research*, *36*(7), 1641–1652. https://doi.org/10.1029/2000WR900092

Palm, C., Blanco-Canqui, H., DeClerck, F., Gatere, L., & Grace, P. (2014). Conservation agriculture and ecosystem services: An overview. *Agriculture, Ecosystems & Environment*, *187*, 87–105. https://doi.org/10.1016/J.AGEE.2013.10.010

Peña-Sancho, C., López, M. V., Gracia, R., & Moret-Fernández, D. (2016). Effects of tillage on the soil water retention curve during a fallow period of a semiarid dryland. *Soil Research*, *55*(2), 114. https://doi.org/10.1071/SR15305

Pires, L. F., Borges, J. A. R., Rosa, J. A., Cooper, M., Heck, R. J., Passoni, S., & Roque, W. L. (2017). Soil structure changes induced by tillage systems. *Soil and Tillage Research*, *165*, 66–79. https://doi.org/10.1016/j.still.2016.07.010

R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rasmussen, K. J. (1999). Impact of ploughless soil tillage on yield and soil quality: A Scandinavian review. *Soil and Tillage Research*, *53*(1), 3–14. https://doi.org/10.1016/S0167-1987(99)00072-0

Reicosky, D. C., & Allmaras, R. R. (2003). Advances in Tillage Research in North American Cropping Systems. *Journal of Crop Production*, *8*(1–2), 75–125. https://doi.org/10.1300/J144v08n01_05

Reicosky, D. C., & Forcella, F. (1998). Cover crop and soil quality interactions in

agroecosystems. *J. Soil Water Conserv.*, *53*, 224–229.

Sanz-Cobena, A., Lassaletta, L., Aguilera, E., Prado, A. del, Garnier, J., Billen, G., … Smith, P. (2017). Strategies for greenhouse gas emissions mitigation in Mediterranean agriculture: A review. *Agriculture, Ecosystems & Environment*, *238*, 5–24. https://doi.org/10.1016/J.AGEE.2016.09.038

Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., … Trumbore, S. E. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, *478*(7367), 49–56. https://doi.org/Doi 10.1038/Nature10386

Schwen, A., Bodner, G., Scholl, P., Buchan, G. D., & Loiskandl, W. (2011). Temporal dynamics of soil hydraulic properties and the water-conducting porosity under different tillage. *Soil and Tillage Research*, *113*(2), 89–98. https://doi.org/10.1016/j.still.2011.02.005

Shelton, D., Jasa, P., Brown, L., & Hirschi, M. (2000). Water Erosion. In I. Ames (Ed.), *Conservation Tillage Systems and Management* (2nd ed.). MidWest Plan Service, Iowa State University. MWPS-45.

Simunek, J., Sejna, M., Saito, H., Sakai, M., & van Genuchten, M. T. (2013). The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media, Version 4.17. Riverside, USA: University of California, Riverside.

Strudley, M. W., Green, T. R., & Ascough, J. C. (2008). Tillage effects on soil hydraulic properties in space and time: State of the science. *Soil and Tillage Research*, *99*(1), 4–48. https://doi.org/10.1016/J.STILL.2008.01.007

Tavares Filho, J., & Tessier, D. (2009). Characterization of soil structure and porosity under long-term conventional tillage and no-tillage systems. *Revista Brasileira de Ciência Do Solo*, *33*(6), 1837–1844. https://doi.org/10.1590/S0100-06832009000600032

Twarakavi, N. K. C., Sakai, M., & Simunek, J. (2009). An objective analysis of the dynamic nature of field capacity. *Water Resources Research*, *45*(10), 1–9. https://doi.org/10.1029/2009WR007944

Upadhyaya, S. K., Lancas, K. P., Santos-Filho, A. G., & Raghuwanshi, N. S. (2001). One-pass tillage equipment outstrips conventional tillage method. *California Agriculture*, *55*(5), 44–47. https://doi.org/10.3733/ca.v055n05p44

Veenstra, J. J., Horwath, W. R., & Mitchell, J. P. (2007). Tillage and Cover Cropping Effects on Aggregate-Protected Carbon in Cotton and Tomato. *Soil Science Society of America Journal*, *71*(2), 362. https://doi.org/10.2136/sssaj2006.0229

Veenstra, J. J., Horwath, W. R., Mitchell, J. P., & Munk, D. S. (2006). Conservation tillage and cover cropping influence soil properties in SanJoaquin Valley cotton-

tomato crop. *California Agriculture*, *60*(3), 146–153.

Veihmeyer, F. J., & Hendrickson, A. H. (1931). The moisture equivalent as a measure of the field capacity of soils. *SOIL SCIENCE*, *32*(3), 181–193. https://doi.org/10.1097/00010694-193109000-00003

Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., … Young, I. M. (2016). Modeling Soil Processes: Review, Key Challenges, and New Perspectives. *Vadose Zone Journal*, *15*(5), 0. https://doi.org/10.2136/vzj2015.09.0131

Villamil, M. B., Bollero, G. A., Darmody, R. G., Simmons, F. W., & Bullock, D. G. (2006). No-Till Corn/Soybean Systems Including Winter Cover Crops. *Soil Science Society of America Journal*, *70*(6), 1936. https://doi.org/10.2136/sssaj2005.0350

Zuber, S. M., & Villamil, M. B. (2016). Meta-analysis approach to assess effect of tillage on microbial biomass and enzyme activities. *Soil Biology and Biochemistry*, *97*, 176–187. https://doi.org/10.1016/j.soilbio.2016.03.011

Chapter 4 Prediction of Soil Moisture at High Spatial Resolution Using Machine Learning and Unmanned Aircraft Systems-Based Remote Sensing

Abstract

We developed a powerful machine learning model to interpret soil moisture at a high spatial resolution based on multispectral imagery captured by small unmanned aircraft system (UAS). Using photogrammetry from the images, we generated a high resolution (10 cm) digital elevation model (DEM) and calculated several topographic parameters at multiple scales. We identified 48 relevant predictor variables; the top most important variables were cumulative precipitation and potential evapotranspiration, reflectance in the red band, and topographic position indices (TPI). Our results demonstrate that the dynamics of soil water status across heterogeneous terrain may be adequately described and predicted by UAS remote sensing data and machine learning. Our modeling approach and the variable importance and relationships we have assessed in this study should be useful for management and environmental modeling tasks where spatially explicit soil moisture information is essential.

List of Acronyms

ALE	Accumulated local effects
ANN	Artificial neural network
BRT	Boosted regression trees
DEM	Digital elevation model
MAE	Mean absolute error
MBE	Mean bias error
ML	Machine learning
NDVI	Normalized difference vegetation index
NIR	Near-infrared
PET	Potential evapotranspiration
RF	Random forest
RMSE	Root mean square error
RVR	Relevance vector regression
SVR	Support vector regression
TDR	Time-domain reflectometry
TPI	Topographic position index
TTVI	Thiam's Transformed vegetation index
UAS	Unmanned aircraft systems

4.1 Introduction

The relatively small quantity of water stored in the upper layers of soil plays a key role in terrestrial biology, biogeochemistry, and atmospheric water and energy fluxes. More than half of the solar energy absorbed by the land surface is used to evaporate water (Trenberth et al., 2009) and about 60% of terrestrial precipitation is returned to the atmosphere by evapotranspiration (Seneviratne et al., 2010).

In most environments, soil water storage mainly depends on precipitation and evapotranspiration (Hillel, 1998; Rana & Katerji, 2000), but the distribution of water in the soil is also dependent on the soil hydraulic properties, topography, and other environmental and underground conditions. Therefore, there is no straightforward or accurate method to estimate soil water.

It is very difficult, or even impractical, to acquire data on soil water dynamics by direct measurement over large scales, and there is no robust approach to predict it. The scarcity of soil moisture observations is a major impediment for the investigation of soil moisture-climate interaction. New techniques to large-scale measurement of soil moisture include the cosmic-ray soil moisture observing system, COSMOS, and the GPS interferometric reflectometry (GPS-IR) based methods. The COSMOS employs a network of probes across the U.S. that estimate soil moisture by measuring cosmic-ray neutron radiation intensity above the land surface (Zreda et al., 2012). GPS based methods are also able to estimate soil moisture of a few square meters using GPS signal reflected from the soil. For these new techniques, the empirical confirmation of theoretical predictions of variable measurement depth and standardization of procedures still need to be refined (Ochsner et al., 2013). Remote sensing techniques can fill the need for spatial coverage. The availability and accuracy of satellite remote sensing technologies have been steadily increasing over the years. However, satellite remote sensing methods are often constrained by their spatial and temporal resolution and shallow depth of measurement (Nichols et al., 2011).

Remote sensing methods of retrieving soil moisture provide an alternative to conventional methods of soil moisture measurement, which are impractical at large scales. Furthermore, they enable spatially distributed and frequent observations over a large area, which is difficult to achieve using conventional field measurements (Barrett & Petropoulos, 2014; Petropoulos et al., 2015). A critical challenge to current remote sensing methods of retrieving soil moisture is the lack of imagery with optimum spatial resolutions appropriate for field-scale soil moisture studies and the low re-visit frequency of satellites (Barrett & Petropoulos, 2014; Das & Mohanty, 2006). Alternatives based on manned airborne platforms are limited due to their high operational costs. Another significant challenge to remote sensing methods is estimating the root-zone soil moisture using surface observation obtained from remote sensing. (Nichols et al., 2011; Ochsner et al., 2013)

Several remote sensing methods, particularly from spaceborne deployment, have been developed to retrieve soil moisture using optical, thermal infrared, and microwave sensors.

The retrieval of information from measurements performed by remote sensing is based on the principle that changes in the chemical, physical and structural characteristics of a target determine the variations of its electromagnetic response (Schanda, 1986). The task of retrieving information from remote sensing is complicated by several factors. Ali et al. (2015) outline four general challenges of the retrieval problems: (i) the often complex and non-linear relation between remote sensing measurement and target variables of interest; (ii) the ill-posed nature of the retrieval problem in that electromagnetic response of target is typically the result of contributions from multiple target variables and similar electromagnetic responses may be associated with different physical variables; (iii) the mixed contribution of multiple objects represented within elementary resolution cell; and (iv) the influence of external disturbing factors such as noise, radiation components coming from surrounding of the investigated area and the atmosphere.

Soil moisture retrieval from remote sensing has traditionally been addressed based on either empirical approaches or approaches based on an inversion of physical models. More recently, the use of machine learning techniques gained increased attention because of its ability to tackle many of the limitations with the empirical and physical-based models.

The approaches based on physical models are based on the physical description of the mechanisms involving the interaction of electromagnetic radiation and the target variable. A wide variety of analytic electromagnetic models have been proposed in the literature. The thermal inertia approach (Price, 1977) is one such method that is most commonly used for soil moisture retrieval using thermal infrared (wavelengths between 3.5 and 14 μm) observation (Barrett & Petropoulos, 2014; D. Zhang & Zhou, 2016). Many new soil thermal inertia estimation methods continue to be developed (Price, 1985; Tian et al., 2015; D. Zhang & Zhou, 2016). The advantages of such physically based models are that they can operate in more general scenarios that are difficult to represent through the collection of in situ measurements. However, such models rely on simplifying the representation of a real phenomenon, which can reduce reliability. The increased complexities of more through analytical models and the need for a large number of input parameters is a drawback for analytical models (D. Zhang & Zhou, 2016).

Empirical modeling approaches employ statistical regression techniques to develop a mapping function based on couples of in situ measurements of the target variable and corresponding remote sensing measurement (Ali et al., 2015). Water is one of the most significant chromophores in soils and studies have shown that narrow band spectral information in the visible (0.4 – 0.7 μm), near-infrared (0.7 – 1.1 μm) and shortwave infrared (1.1 – 2.5 μm) regions can be used to estimate surface soil moisture (Ben-Dor et al., 2009; Malley et al., 2004). Soil reflectance in the visible to shortwave infrared spectral region generally decreases with increase in soil moisture, with some parts of the spectrum showing more pronounced decrease than others (Haubrock et al., 2008; Weidong et al., 2002). The hydroxide bond is the strongest absorber in the near-infrared region and free water in soil pores has strong absorption around 1.4 and 1.9 μm wavebands (Malley et al., 2004). Several hyperspectral techniques to estimate soil moisture content have been

developed such as the Soil Moisture Gaussian Model (SMGM) (Whiting et al., 2004) and the Normalized Soil Moisture Index (NSMI) (Haubrock et al., 2008).

In the presence of vegetation cover, however, the ability to use soil reflectance to measure soil moisture is limited (Muller & Décamps, 2000). In addition, soil reflectance of solar radiation represents only the upper 50 μm of soil, and this makes it challenging to estimate moisture conditions in deeper layers (Malley et al., 2004). Most soil moisture remote sensing approaches operating in the optical range relay on developing an empirical spectral vegetation index (Barrett & Petropoulos, 2014). Several soil moisture measurement methods based on vegetation index proxies have been suggested as vegetation indexes are extremely sensitive to water stress, and they allow indirect estimates of soil moisture (D. Zhang & Zhou, 2016). Many studies have focused on deriving surface soil moisture content from synergistic use of remote sensing data acquired simultaneously in the optical and thermal infrared spectrum. The so-called 'universal triangular relationship' is a widely used method for estimating soil moisture (Nichols et al., 2011; Sobrino et al., 2014).

The advantage of empirical relationships is that they are typically fast to derive, do not require too many inputs, and have good accuracy (Ali et al., 2015). The disadvantages of empirical models are the need for good quality ground measurement, which could be time consuming and expensive, and that the derived relationship is typically site and sensor dependent which limits the possibility to extend their use in a different area readily.

Some disadvantages specific to remote sensing methods in the optical and thermal infrared spectrum are the fact that these wavelengths have shallow soil penetration and require cloud-free conditions. Many of the optical and thermal infrared synergistic approaches require a wide range of both vegetation index and soil moisture conditions within a study region which cannot always be satisfied (Barrett & Petropoulos, 2014).

Advantages of the machine learning techniques in remote sensing are their ability to learn and approximate complex non-linear mappings and the fact that no assumptions need to be made about data distribution. They can thus integrate data from different sources with poorly-defined or unknown probability density functions (Ali et al., 2015). Machine learning techniques have often been shown to outperform other parametric approaches (Ali et al., 2015; Paloscia et al., 2008). Furthermore, machine learning techniques improve with an increasing number of observed datasets. Some of the limitations of machine learning methods are the need for a large number of training data which require extensive ground truth datasets, and that machine learning methods are black boxes and only limited inference can be made about the relationships of different inputs.

Remote sensing from unmanned aircraft systems (UAS) has the potential to address several limitations of traditional remote sensing. The most attractive feature of UASs is their high spatial resolution, frequent or on-demand image acquisition, and low operating costs (Anderson & Gaston, 2013; Berni et al., 2009; Colomina & Molina, 2014; Elarab, 2016). UAS is an umbrella term that refers to the unmanned aircraft and the complementary

ground control and communication systems necessary for air surveys (Singh & Frazier, 2018).

4.1.1 Objectives

The purpose of this research was to advance soil moisture change measurement, process understanding, and prediction using remote sensing products from UAS and machine learning methods. In this study, the spatial and temporal scale limitations were addressed by deploying multispectral remote sensing with small UAS and address the challenge of retrieving surface soil moisture changes using machine learning methods and fusing remote sensing data with ground data and meteorological data.

The specific goals of this study were to: (1) develop an adaptable method to retrieve information on surface soil moisture from small UAS remote sensing products and machine learning methods, (2) identify important reflectance and surface characteristics for the prediction of soil moisture changes (3) identify appropriate spatial resolutions of reflectance images and terrain variables for estimating soil moisture, and (4) explore the relation of soil moisture to surface properties.

4.2 Background on The Machine Learning Algorithms Used

Numerous ML algorithms exist for multivariate regression modeling. Most commonly used machine learning technique for soil moisture retrieval from remote sensing is the artificial neural network (ANN) (e.g., Hassan-Esfahani et al., 2015; Paloscia et al., 2008). In recent years, the support vector machine (SVM) based support vector regression (SVR) regression has become popular in the retrieval of soil moisture (e.g., Ahmad, Kalra, & Stephen, 2010; Zaman & Mckee, 2014; Zaman, McKee, & Neale, 2012). Other popular machine learning algorithms include tree-based models such as the Random Forest (RF) and Boosted Regression Trees (BRT).

4.2.1 Artificial Neural Network (ANN)

ANN models have been widely used in the development of PTFs (Matei et al., 2017; Pachepsky et al., 1996; Schaap et al., 2001; Y. Zhang & Schaap, 2017; Y. Zhang et al., 2018). ANNs are universal approximators that can approximate any nonlinear mapping. The feed-forward neural network is a popular variant of ANN. In this study, we implemented the feed-forward neural networks with a single hidden layer.

4.2.2 Support Vector Regression (SVR)

SVR is an adaptation of the support vector machine (SVM) for regression problems (Cortes & Vapnik, 1995; Drucker et al., 1997). The SVM learning is a generalization of 'maximal margin classifier.' The algorithm first maps the input variables into a high-dimensional space using a fixed mapping function—a kernel function. The algorithm then constructs hyperplanes, which can be used for classification or, in the case of SVR, for regression. In

this study, we use the Radial Basis Function kernel, which is one of the most commonly used kernels in SVR. Some advantages of SVR include the fact that they do not suffer from the problem of local minima, and that they have few parameters to tune when training the model.

4.2.3 Relevance Vector Regression (RVR)

Like SVM, the RVR originally introduced as a classification machine (Tipping, 2000). RVR is a Bayesian treatment of the SVM prediction function which avoids some of the limitations of SVM algorithms, such as reducing the use of basis functions and the need for optimizing the cost and the insensitivity parameters (Ben-Shimon & Shmilovici, 2006). Torres-Rua et al. (2016) successfully used the RVR algorithm to estimate surface soil moisture from satellite image and energy balance products.

4.2.4 Random Forest (RF)

RF are popular models that are relatively simple to train and tune (Hastie et al., 2009). They apply ensemble techniques by averaging a large number of individual decision tree-based models. Tree models are 'grown' by searching for a predictor that ensures the best split that results in the smallest model error. The individual trees in RF ensemble are built on bootstrapped training sample, and only a small group of predictor variables are considered at each split, this ensures that trees are de-correlated with each other (Breiman, 2001; James et al., 2013).

4.2.5 Boosted Regression Trees (BRT)

BRT is another form decision tree model ensemble enhanced by the gradient boosting approach. The gradient boosting algorithm constructs additive regression models by sequentially fitting 'simple base learner' functions (i.e., decision trees) to current pseudo-residuals at each iteration (Friedman, 2002). These pseudo-residuals are the gradient of the loss function being minimized. BRT models have shown considerable success and often outperform other ML algorithms in many situations (Elith et al., 2008; Natekin & Knoll, 2013). BRT models are also particularly adept for less-than-clean data (Friedman, 2001), which makes them particularly attractive in our work where the training data is compiled from various sources and different measurement methods which makes it prone to some inconsistencies.

Tree-based models, both the RF and BRT, have the advantage of being able to rank predictor variable's relative importance. In these models, the approximate relative influence of a single predictor variable is calculated as the empirical improvement of predictions by splitting on that predictor at each node and then averaging the relative influence of the variable across all trees of the model (Ridgeway, 2012).

4.3 Methods

Multispectral images of the study area were collected on six different days throughout the 2018 water year using a UAS equipped with a multispectral camera. High-resolution digital elevation model (DEM) was generated from the stereo images using photogrammetric software, and multiple sets of terrain variables were calculated. Concurrently with the

image acquisition flights, moisture content of top 3.8 cm of soil was measured at predefined sampling locations. The ground soil moisture measurements, multispectral reflectance, terrain variables, and rainfall and potential evapotranspiration (PET) data were then aggregated into a data table and used to train a machine learning model to predict the soil moisture. Figure 4-1 shows the model building process.

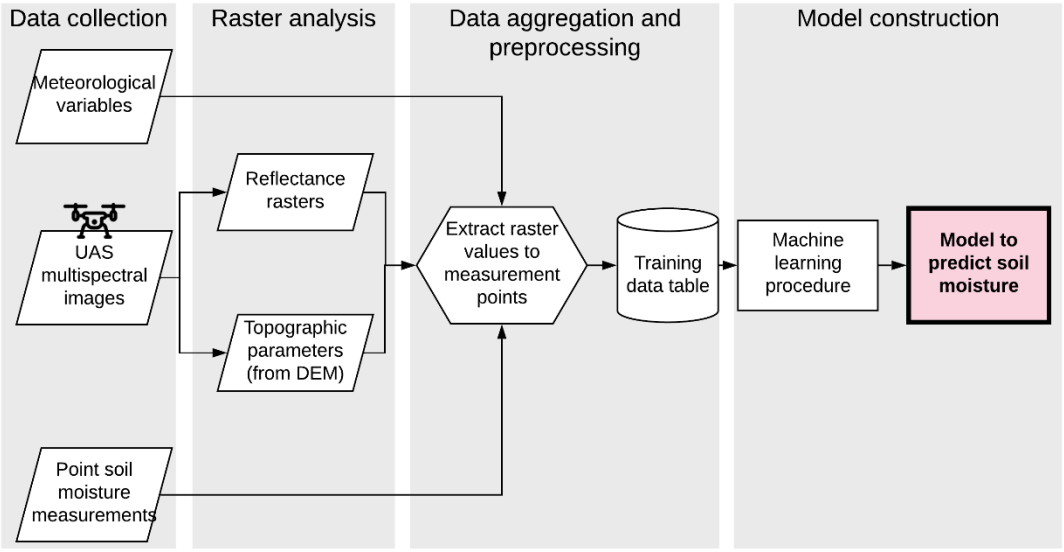

Figure 4-1 Process flowchart of model development.

The methods section is divided into five subsections. The first subsection describes the study site. The second subsection describes how data was collected from the three main sources: ground sampling, UAS imagery, and meteorological data. The third subsection discusses data processing and how new variables are calculated. The fourth subsection details the core machine learning procedure followed to construct the model. Finally, the fifth subsection describes the procedure for using the machine learning models to predicts soil moisture over the landscape.

4.3.1 Study Site

The study was conducted in a small grassland catchment at the Merced Vernal Pools and Grassland Reserve located about five kilometers northeast of the City of Merced, California. Located at the Central Valley of California, the study site has a Mediterranean climate with hot, dry summers and cool, wet winters with an average annual precipitation of 330 mm.

The Merced Vernal Pools and Grassland reserve covers an area of about 26.6 km^2 and protects hundreds of ephemeral pools and wetlands (Wong, 2014). The reserve was historically and still is used for livestock grazing.

Our study site is a 0.6 km^2 area of land located within a sub-catchment that contribute to the Avocet Pond, a large stock pond located in the northeast corner of the Reserve (Figure 4-2). The catchment was selected because of an extensive hydrologic modeling study that was being conducted on the site at the time (Fryjoff-Hung, 2018). We studied a small subset of an area investigated by Fryjoff-Hung (2018) to ensure similar land properties represented by our ground sampling.

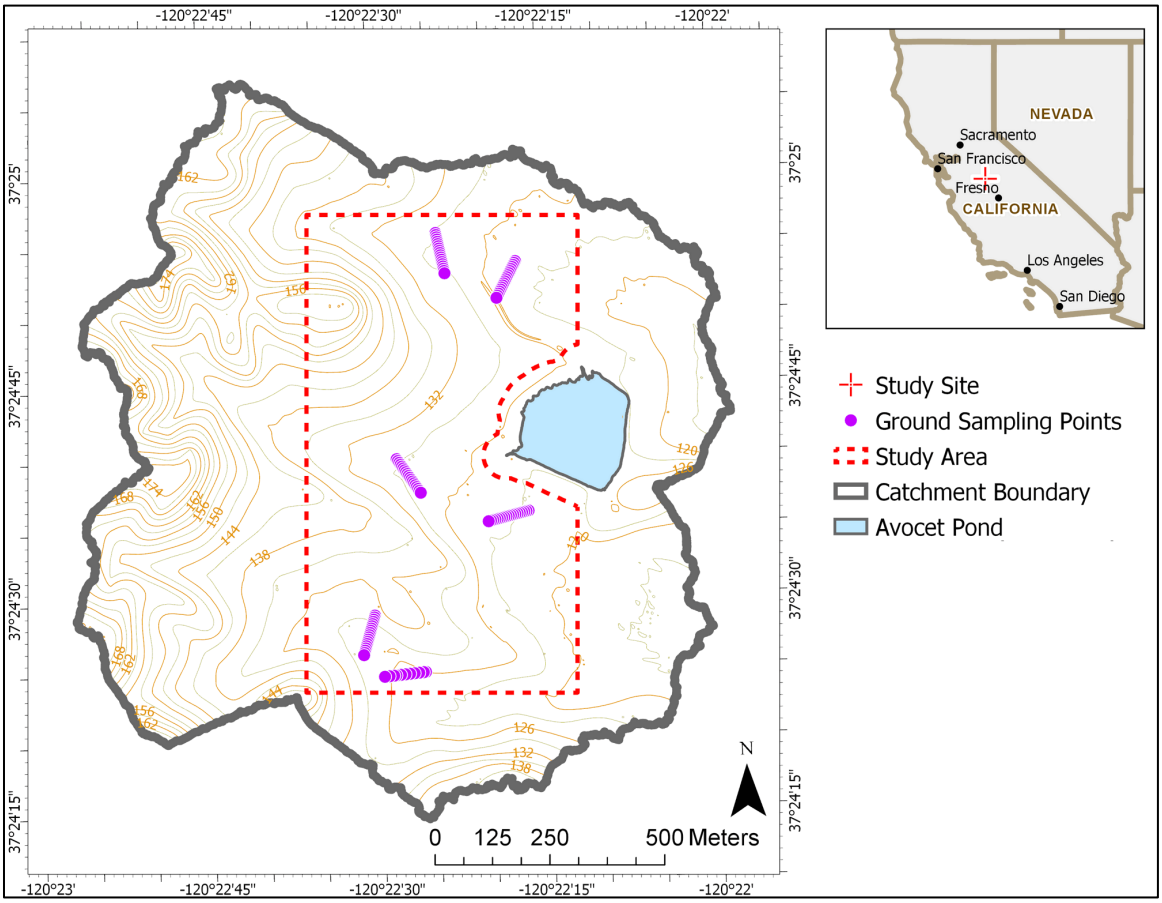

Figure 4-2 Map of Avocet Pond catchment showing the footprint of the study area, ground sampling points, and elevation contours in meters. Inset shows the location of the study site (red crosshair) in California.

The study area soils are dominated by Redding gravelly loam (Fine, mixed, active, thermic Abruptic Durixeralfs) soils. The elevation of the study area ranges from 118 to 162 m above sea level, and slope ranges from 0 to 31°. The distribution of four topographic variables within the study site is given in Figure 4-3.

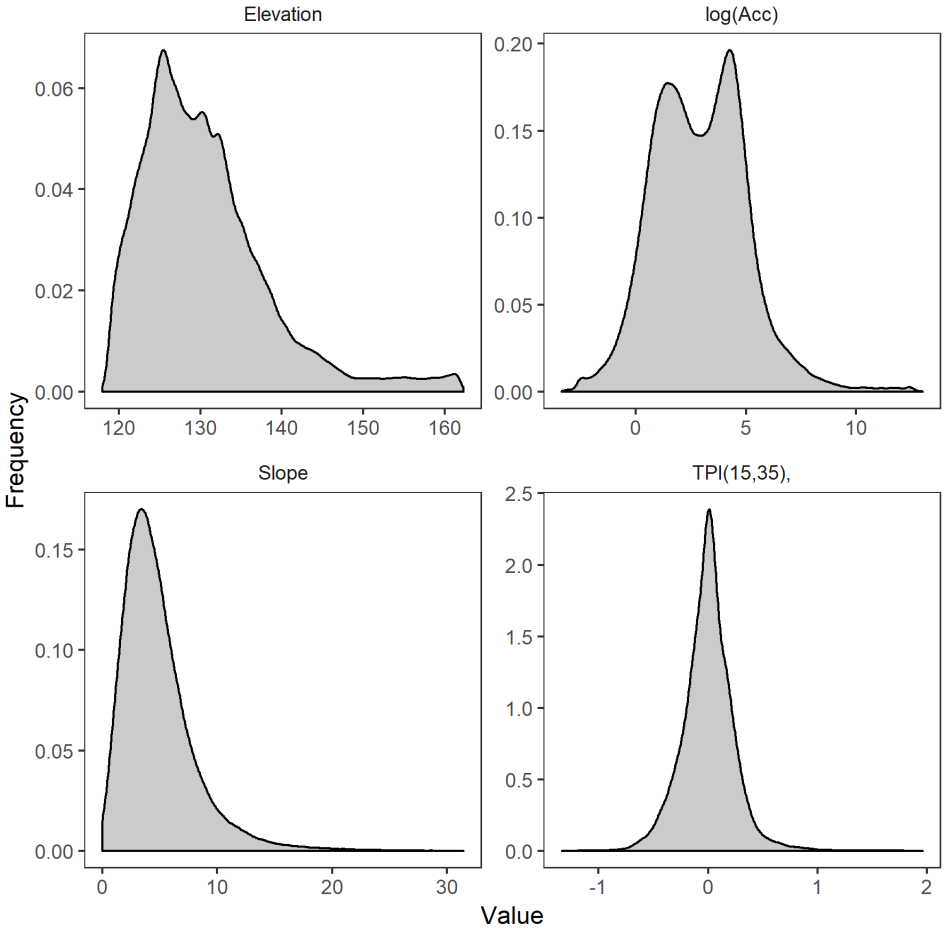

Figure 4-3 Kernel density estimate of values from four topographic variables for the study site. Elevation, flow accumulation, and slope variables were derived from 1 m resolution raster. (See Table 4-1 for a description of the variables)

The fact that vernal pool ecology is predominantly controlled by large seasonal shifts and high spatial variability in hydrology make the study site particularly attractive to the proposed research. UAS have the potential to provide information at appropriate spatial and temporal scales for vernal pool studies (Stark et al., 2015). Knowing soil moisture dynamics at a higher frequency is important, especially during the seasonal transition times. The annual seasonal cycle of the study site is shown in Figure 4-4.

Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Transition to wet		Wet				Transition to dry		Dry			

Figure 4-4 Vernal pool annual moisture cycle.

4.3.2 Data Collection

The imagery was acquired on six days during the 2018 water year green-up and brown-down using a fixed-wing UAS with a multispectral camera onboard (See Table 4-3). Figure 4-5 shows a typical scene of the study site during wet and dry seasons. Point soil moisture measurements (top 4 cm) were collected with a time-domain reflectometry (TDR) probe across precise sampling transects identified with real-time kinematic (RTK) positioning survey. Daily rainfall and PET values were acquired from nearby weather stations.

Figure 4-5 Typical scene of the study area on April (left) and June(right) of 2018.

4.3.2.1 Image acquisition and processing

The UAS remote sensing flights were conducted in the late mornings to mid-day during clear weather conditions. A single remote sensing campaign takes approximately three to four hours, meaning images were acquired between approximately 10:00 AM and 2:00 PM.

Multispectral images were acquired using *Parrot Sequoia* Sensor (Parrot SA, Paris, France) equipped with a sunshine sensor that measured irradiance at the sensor spectral wavebands for radiometric normalization. The camera is deployed on a fixed-wing unmanned aircraft (Finwing Sabre, Finwing Technology) with an average flight height of 120 m above ground level. Images of a calibrated reflectance panel (MicaSense, Inc, Seattle, WA) were taken before each flight and used in the radiometric calibration of images.

The Parrot Sequoia sensor captures four separable bands in the green, red, red edge and near-infrared bands with a focal length of 3.98 mm and resolution of 1280x960 pixels. A fifth channel captures a high-resolution image in the visible spectrum with a focal length of 4.88 mm and resolution of 4608x3456 (Pix4D, n.d.). About 12,000 images are captured per flight with a ground pixel resolution of approximately 10 to 15 cm. Images are

mosaicked, orthorectified, and radiometrically calibrated using Pix4D photogrammetry software (Pix4D, Lausanne, Switzerland). The Pix4D software also generates DEM photogrammetrically from stereo-images.

4.3.2.2 In situ soil moisture measurement

The moisture content of the top 4 cm soil was measured simultaneously with UAS remote sensing flights using FieldScout TDR-300 soil moisture meter equipped with a 3.8 cm probe (Spectrum Technologies Inc., IL, USA). The FieldScout TDR-300 measures volumetric water content using time-domain reflectometry with a resolution of 0.1% and an accuracy of $\pm 3\%$ (Spectrum Technologies Inc., 2009).

We identified six, 90 m long transects within the watershed as sampling transects. To ensure that the sampling transects run over a variety of topography and do not fall within a topographically homogenous area, we generated hydraulic terrain variables using the DEM prior to selection of transect locations. We generated sub-basin boundaries, flow accumulation, topographic wetness index, and stream network maps. We laid out the sampling transects in a way that they traversed across multiple values in terms of topographic wetness index and flow accumulation and ensuring that they cross, and not follow, a stream network. We placed the six transect so that each fell in a separate sub-basin within the Avocet basin. Three of the transects fall in separate sub-basins that feed into the Avocet pond, and the remaining three are in sub-basins below the Avocet pond.

Once we decided on the location of the sampling transects the location of the two ends was recorded accurately using RTK positioning survey and marked with a metal peg hammered into the ground to allow for repeated measurement at the same location. During the soil moisture measurement campaign, we temporarily affixed a 90 m tape measure at the two ends of the transect and took soil moisture measurement about every 10 m noting the exact distance of the sampling point from the transect ends.

4.3.2.3 Meteorological variables

Daily precipitation data was retrieved from the UC Merced weather station located approximately six km southwest of the study site (California Department of Water Resources, 2018). Daily reference evapotranspiration data was retrieved from the California Irrigation Management Information System's Merced station located approximately 10 km south of the study site (California Irrigation Management Information System, 2018). The Reference evapotranspiration is evapotranspiration from standardized grass calculated using the modified Penman (CIMIS Penman) and the Penman-Monteith equations (California Irrigation Management Information System, n.d.). The reference evapotranspiration is considered as PET in this study.

4.3.3 Data Processing

To prepare the data for machine learning. We compiled all the information into a table with the measured soil moisture from each sampling point and date organized into one column. Each row contained the accompanying information for that sampling point and time.

4.3.3.1 Multispectral image pre-processing

4.3.3.1.1 Geometric Correction

Pix4D software was used to process the images. Between 7 to 9 ground control point targets (GCPs) with precise locations identified by RTK survey were used for photo alignment. The mean georeferencing root-mean-square-errors (RMSE) of the GCPs ranged from 0.6 – 2 cm, and mean reprojection errors ranged from 0.1 – 0.2 pixel based on the bundle block adjustment error assessment report. DEM was generated using the structure-from-motion technique; noise filtering and mild surface smoothing (sharp smoothing) were applied to correct for noisy and erroneous points of the point cloud. The inverse distance weighting algorithm was used to interpolate between points to create the raster DEM.

4.3.3.1.2 Radiometric correction

Along with capturing images in the four spectral bands, the multispectral camera records the location, orientation, and solar irradiation using its GPS, inertial measurement unit (IMU), and sunshine sensor, respectively. Radiometric calibration by the Pix4D software considers the positional data, solar irradiance measurements, and gain and exposure data from the camera to convert raw digital numbers into sensor reflectance values. Sensor reflectance represents the ratio of the reflected light to the incoming solar radiation and provides a standardized measure which is directly comparable between images. Finally, surface reflectance is calculated in post-processing, taking into account the camera's orientation, the angle of the sun, and the known reflectance values of the calibration panel.

4.3.3.2 Feature engineering

We calculated several variables based on the multispectral reflectance, terrain, and meteorological data to be used to train a machine learning model as predictor variables. A list of all the measured and calculated variables used in modeling soil moisture are given in Table 4-1.

4.3.3.2.1 Reflectance based vegetation index

We calculated the Thiam's Transformed vegetation index (TTVI) based on the red and near-infrared bands (Equation 3.1).

$$TTVI = \sqrt{\left|\frac{NIR - R}{NIR + R} + 0.5\right|} \tag{3.1}$$

The TTVI is a transformation of the commonly used Normalized Difference vegetation index (NDVI). The reason for choosing of TTVI over NDVI is that it eliminates negative values and transforms NDVI histograms into a normal distribution.

4.3.3.2.2 Terrain variables

A list of the calculated topographic variables and their description is given in Table 4-1. Topographic variables derived from DEM are scale-dependent, to account for this, we calculated all topographic variables on six different resolution DEM. For this, we first

upscaled the DEM from the original resolution of 6.85 cm to 15, 30, 60, 100, 300, and 500 cm cell resolution. We then calculated topographic variables on all the resolutions.

The calculation of topographic position index (TPI) does not only depend on DEM resolution but on the definition of inner and out radii of the annulus (see Equation 3.2).

$$TPI = Elevation - focal\ mean\big(annulus(Inner\ Radius, Outer\ Radius)\big) \quad (3.2)$$

We calculated TPI for different neighborhood sizes using the ArcGIS 10.5 Land Facet Corridor Tool (Jenness et al., 2013). We calculated TPI on three DEM resolutions (100, 300 and 500 cm) with two inner radii (1 and 3 cells) and three outer radii (3, 5, and 7 cells).

The raster images of selected topographic variables are shown in Figure 4-6.

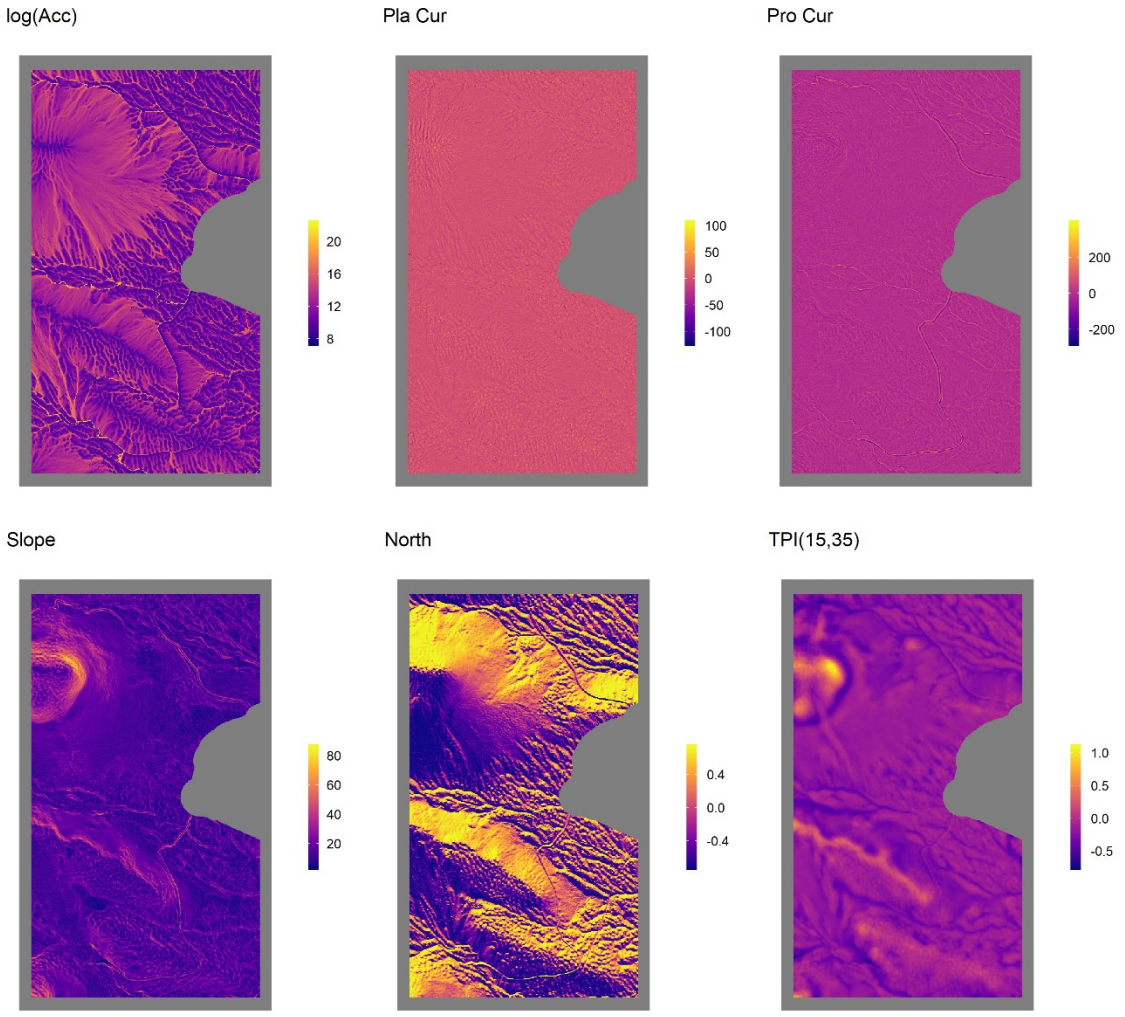

Figure 4-6 Values of some, 1 m resolution, topographic variables over the study area.

4.3.3.2.3 Meteorological variables

Precipitation and PET are two important drivers of surface soil moisture. We used the cumulative water year precipitation and PET. We also calculated rolling sums of those variables across different time span before the measurement, we calculated 1-, 2-, 3-, 7-, 15, and 30-day cumulative precipitation and PET before sampling dates and used those rolling sums as input.

Table 4-1 Measured and calculated data used for machine learning. All topographic variables are computed from the digital elevation model. Descriptions and significance of topographic variables adapted from Wilson & Gallant (2000).

	Variable [unit]	Description	Significance/ relation to soil moisture
Measured	Soil moisture content [%]	Volumetric soil moisture content	Variable of interest.
	Daily rainfall [mm]	Daily rainfall from Precipitation Gage (OTT Pluvio) with a windshield	Source of soil moisture
	Green [-]	Surface reflectance in the green wavelength band (530 – 570 nm)	Soil and vegetation reflectance change
	Red [-]	Surface reflectance in the red wavelength band (640 – 680 nm)	Soil and vegetation reflectance change
	Red-edge [-]	Surface reflectance in the red-edge wavelength band (730 – 740 nm)	Soil and vegetation reflectance change
	Near-infrared [-]	Surface reflectance in the near-infrared wavelength band (770 – 810 nm)	Soil and vegetation reflectance change
	Altitude [m]	Elevation (m)	Vegetation, potential energy
Calculated	Daily potential evapotranspiration [mm]	Reference evapotranspiration from standardized grass calculated using CIMIS Penman equation.	Major soil moisture loss pathway.
	Thiam's Transformed Vegetation Index (TTVI) [-]	$TTVI = \sqrt{\left\lvert \left(\frac{NIR - R}{NIR + R}\right) + 0.5 \right\rvert}$	Vegetation moisture stress.
	Slope [degrees]	Slope gradient (degrees)	Surface and subsurface flow velocity, runoff rate, vegetation, geomorphology
	Aspect [$\cos(degees)$]	Cosine transformed direction of maximum downward gradient (northerness).	North and south-facing slopes differ in solar insolation, PET, flora and fauna distribution, and abundance.
	Profile curvature [-]	Downslope curvature	Flow acceleration, erosion/deposition rate, geomorphology
	Plan curvature[-]	Along-side curvature	Converging/diverging flow, soil characteristics
	Tangential curvature[-]	Curvature in an inclined plane	Represents areas of convergent (concave) and divergent (convex) flow.
	Flow accumulation (MFD methods) (A) [cm^2]	Catchment area draining to pixel	Runoff volume, geomorphology
	Length-Slope factor, LS [-]	Length-slope factor from the Revised Universal Soil Loss Equation (RUSLE). For slope lengths <100 m and slopes <14°: $LS = 1.4\left(\frac{A}{22.12}\right)^{0.4}\left(\sin\frac{S}{0.0896}\right)^{1.3}$	Calculates a spatially distributed sediment transport capacity
	Topographic position index, based SFD and MFD methods [-]	$TPI = Z_0 - \frac{1}{n_R}\sum_{i \in R} Z_i.$	
	Topographic wetness index, based on SFD and MFD methods [-]	$TWI = \ln\left(\frac{A}{\tan S}\right)$	Commonly used index to quantify topographic control on hydrological process.

4.3.3.3 Data aggregation

We had a total of 406 soil moisture measurements across the six different measurement times. For each soil moisture, we extracted the raster values of reflectance and terrain variable by using raster to point data extraction tool in ArcGIS Pro using a 100 cm radius means. For this, ground coordinates of sampling locations were overlaid onto the geo-referenced image, and the pixel values across the bands representing the center of each sampled area were extracted to result in tabular data. We also included the meteorological variables for each soil moisture reading.

4.3.3.3.1 Variable transformation

Prior to modeling, all the predictor variables were standardized by centering the training variable's mean to zero and scaling by the variable's standard deviation, as shown in Equation (3.3):

$$x' = \frac{x - \bar{x}}{\sigma_x} \tag{3.3}$$

where x' is the centered and scaled value of variable x and; $\bar{x}$ and σ_x are the arithmetic mean and standard deviation of the variable.

4.3.3.3.2 Variable selection

Variable selection (or feature selections) involves the selection of a subset of relevant variables (features) from a larger set of potential predictors. Benefits of variable selection include improvement of model performance, reducing training and utilization times, and facilitating data understanding (Guyon & Elisseeff, 2003; Weston et al., 2003). We employed three methods of variable selection: test of linear correlation and linear dependencies among variables, and recursive feature elimination. Recursive feature elimination involves removing the least important features whose omission has the least effect on training errors (X. Chen & Jeong, 2007; Guyon et al., 2002). We implemented recursive feature elimination procedure during the coarse tuning of BRT, RF, ANN, and SVR algorithm models.

The data preparation resulted in 138 variables. Of these, 76 variables were removed based on linear correlation and linear dependencies among variables. An additional 16 were removed following the recursive feature elimination procedure. The final data used for building the models had 46 variables (Table 4-2), of which five are meteoric, nine are reflectance variables, and 32 are topographic variables. Variable categories that had no importance included the topographic wetness index (TWI), the reflectance in the red-edge band, and NDVI.

Table 4-2 Important predictor variables used in final models.

Domain	Variable	Scale*
Meteoric	Potential evapotranspiration	1, 30
	Precipitation	1, 15, 30
Reflectance	Green	0.6, 1, 3
	Red	0.6, 1, 3
	Near Infrared	0.6, 1, 3
Topographic	Northerness	0.6, 1, 3, 5
	Slope	0.6, 1, 3, 5
	Flow Direction	0.6, 1, 3, 5
	Flow Accumulation	0.6, 1, 3, 5
	Curvature (Profile)	1, 3, 5, 50
	Curvature (Planform)	0.6, 1, 3, 5, 50
	Topographic Position Index	(1,3), (3,7), (3,9), (5,15), (9,21), (15,35), (15,100)

* Scale for raster products is pixel resolution in meters, and cumulative days for the meteoric variables. Topographic Position Index scale is a combination of the inner-outer diameters in meters.

4.3.4 Data Description

The data collection days and summary site statistics are given in Table 4-3. Measurements were done during six days throughout the 2018 water year.

Table 4-3 Data collection days and site summary statistics.

	Date	Day of the water year	Cumulative water-year precipitation [mm]	Cumulative water-year PET [mm]	Mean soil moisture (and standard deviation) [%]	Sample count
1	2017-10-30	30	5.8	96.58	2.47 (1.09)	60
2	2017-12-04	65	34.3	149.16	12.26 (5.23)	60
3	2018-01-23	115	85.6	199.15	24.85 (7.79)	64
4	2018-04-04	186	133.3	371.81	20.27 (10.72)	92
5	2018-05-01	213	177.9	489.18	8.98 (3.93)	74
6	2018-05-24	236	177.9	619.26	4.42 (1.68)	56

The cumulative precipitation and cumulative PET for the 2018 water year are shown in Figure 4-7. The 1, 2, 3, 7, 15, and 30-day rolling sums of precipitation and PET for the six sampling dates from days are shown in Figure 4-8.

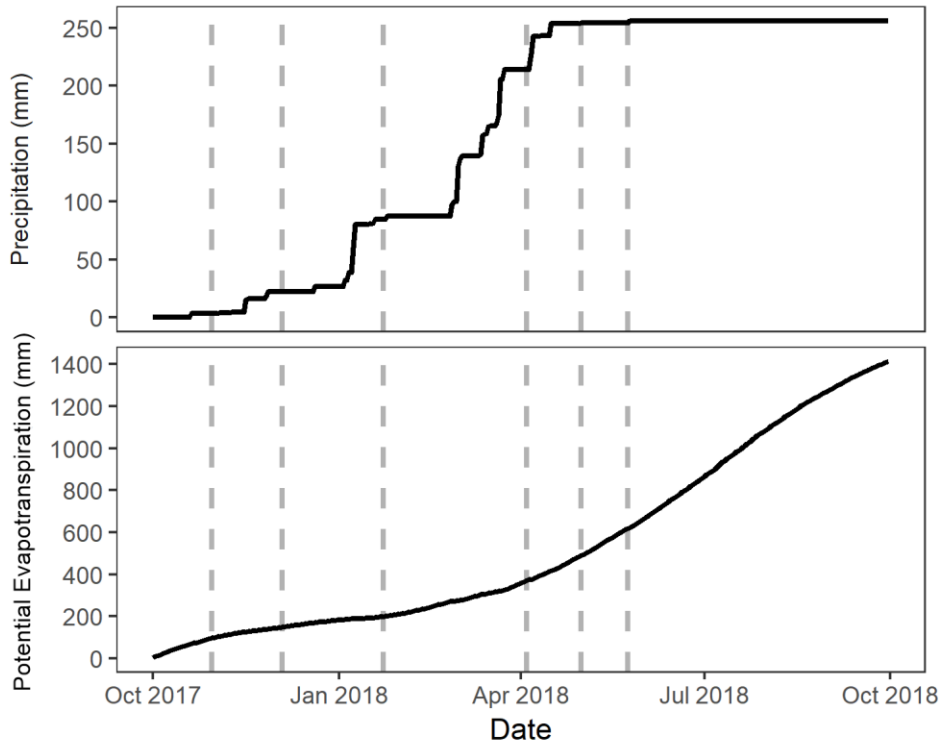

Figure 4-7 Cumulative precipitation and potential evapotranspiration for the 2018 water year (Source: UC Merced, and CIMIS Merced stations, respectively). Vertical dashed lines indicate measurement dates.

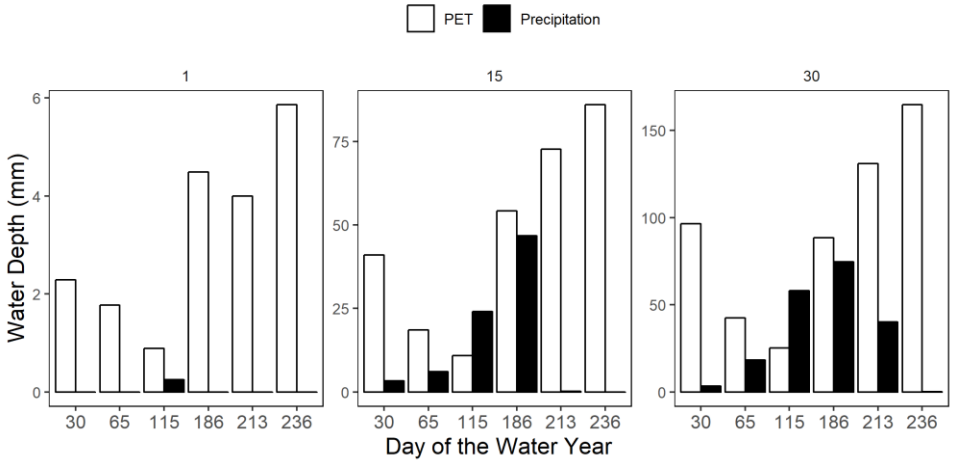

Figure 4-8 Rolling sums of precipitation and potential evapotranspiration by window size in days before the sampling date.

The soil moisture measurement followed the general precipitation patterns but was also influenced by immediate rainfall events; the highest soil moisture occurred on the only measurement day where it had rained the day before (January 23, 2018). The vegetation greenness, as measured with TTVI followed the 15-day cumulative rainfall well with

maximum greenness occurring on April 4, 2018, and sharply decreasing the following two months (Figure 4-9).

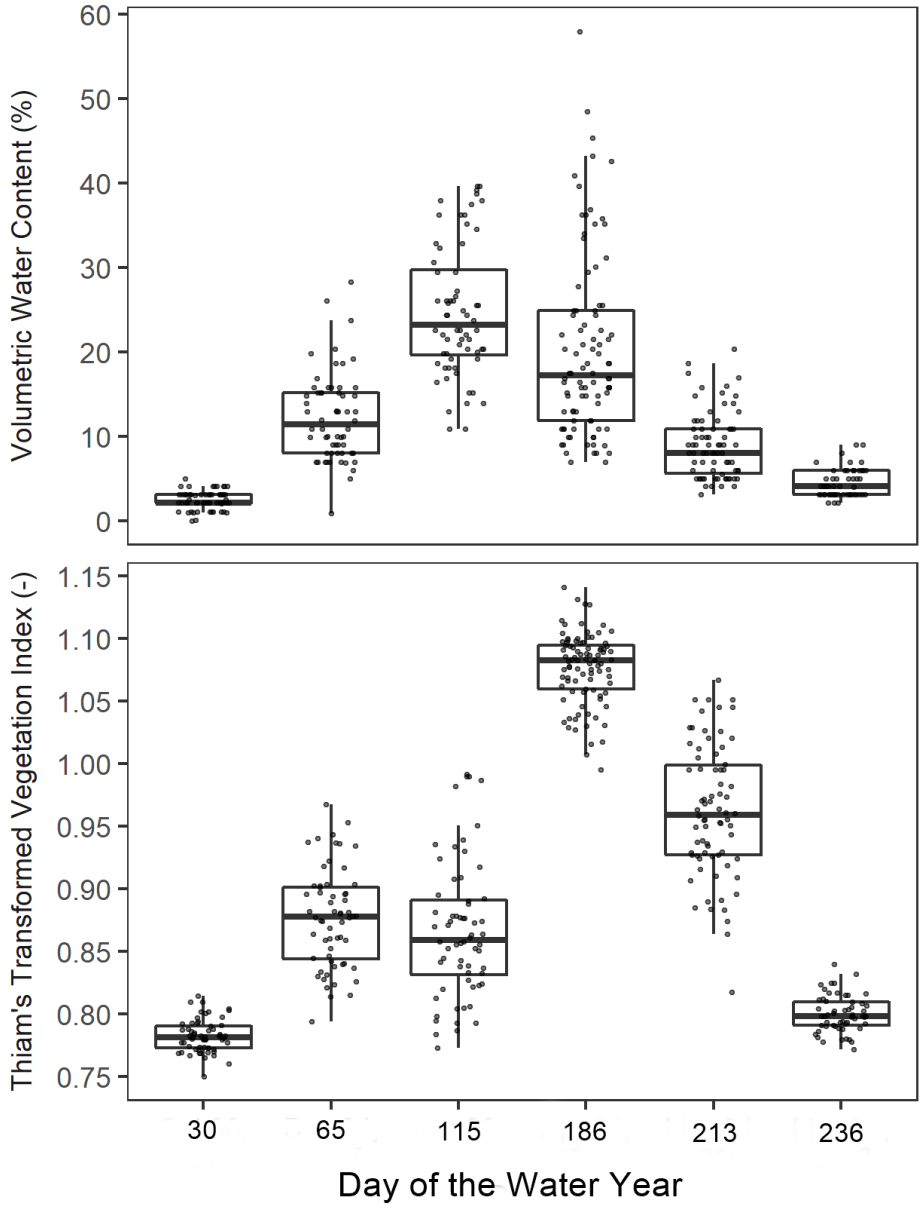

Figure 4-9 Measured soil moisture and vegetation index of the ground sampling locations from 1 m resolution raster.

Appendix C, Figure A0-8 shows the distribution of some terrain variables of the ground sampling points derived from the digital elevation model and correlations between variables. The terrain variables for the ground sampling points show a reasonable distribution in values, while the distribution of elevation shows a bimodal distribution with ranges from 120 to 130 m, most of the other variables show a close to normal distribution.

The only variables with Person's correlation above 0.5 are between TPI and curvature (Pearson's correlation = 0.67). The distribution of values for the variables selected variables in the data is shown in Appendix C, Figure A0-9.

4.3.5 Machine Learning Procedure

The overall procedure of building the ML models is illustrated in Figure 4-1. The computationally demanding steps of model training and testing were run at the Multi-Environment Research Computer for Exploration and Discovery (MERCED) high-performance computing cluster, at the University of California, Merced. The *caret* R package (Kuhn, 2017) was used to handle training and tuning procedures. The SVR and RVR algorithms were implemented using the *kernlab* package (Karatzoglou et al., 2004), RF algorithm was implemented using the *randomForest* package (Liaw & Wiener, 2015), and the BRT algorithm using the *xgboost* package (T. Chen & Guestrin, 2016).

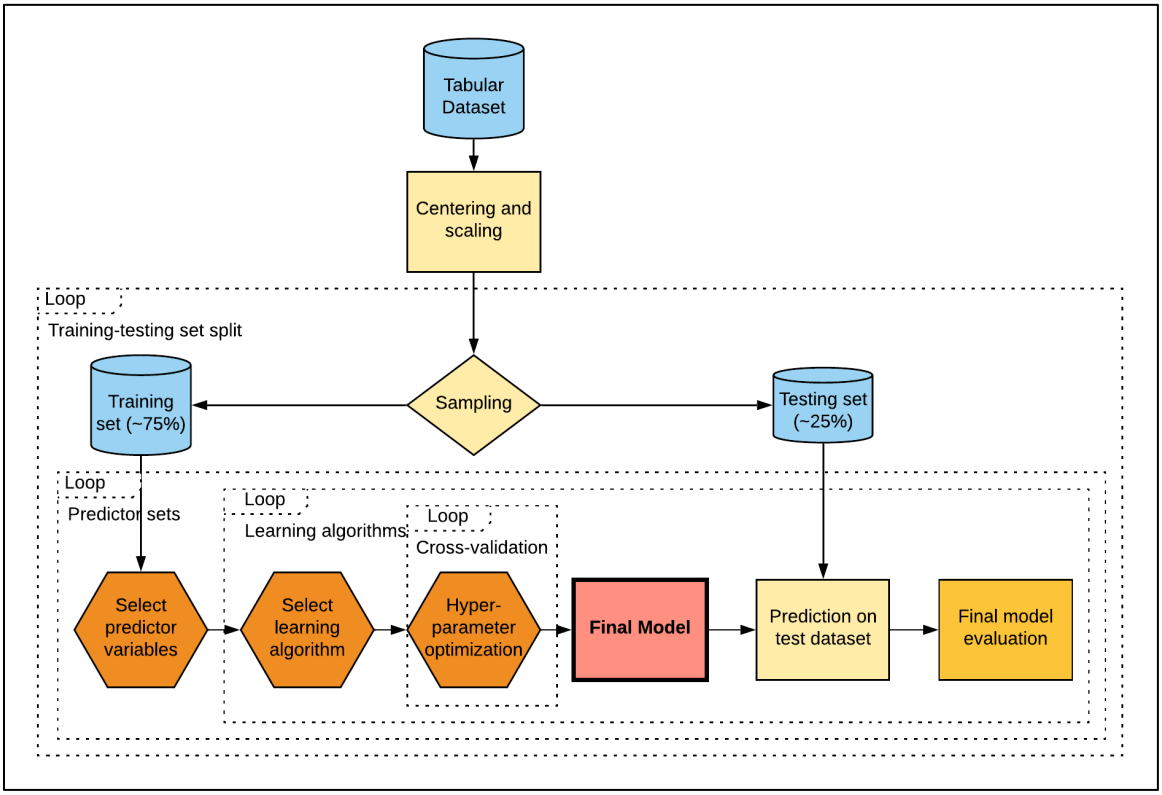

Figure 4-10 Flowchart showing the model training process.

4.3.5.1 Training-testing set splits

The data was split into training and testing sets of approximately 75-25 percent, respectively (i.e., 300 and 100 records). The testing set was a hold-out set used only to evaluate final trained models.

For the testing set, two transects are randomly selected on four randomly selected sampling dates, and one transect is randomly selected on the remaining two sampling dates. To minimize bias that may result from the training-testing set split, we generated 30 unique training-testing set splits and trained 30 separate models based on each separate training set. The performance of each model was assessed on its respective testing set. Similar performance of the individual models would indicate that bias due to the training-testing set split is minimal. The justification for this subsetting procedure is: (1) the selection of entire transects as testing sets avoids the possible data leakage between the training and testing sets due to spatial autocorrelation since samples in a transect are located close to each other—a simple random splitting would not avoid this potential problem; (2) all six sampling dates are represented in the training set—models are trained on the entire range of time and soil moisture changes; and (3) the testing set is between 25 to 30 percent of the data (between 100 to 125 samples).

The distribution of samples across the sampling dates and transects for the training and testing sets are shown in Appendix C, Figure A0-10 and Figure A0-11, respectively.

On average the training-testing split was 294 samples in the training sets and 113 samples in the testing sets. All the training sets have samples from all the six sampling dates and transects. While all sampling dates are represented in each testing set, on average, there are five transects in each testing set.

4.3.5.2 Cross-validation procedure

The selection of optimal model parameters in the model training process was done by the cross-validation method. Cross-validation is done to estimate the test error rate by holding out a subset of the training data (i.e., validation set) from the fitting process and then applying the fitted model to predict the validation subset. A 30-fold cross-validation set was generated by randomly splitting the training data into 80-20 percent training-validation split by randomly selecting a single transect every day. Optimum model parameters were selected using a comprehensive grid search method.

4.3.5.3 Model assessment

4.3.5.3.1 Performance

The final performance of models was assessed on the separate hold-out test dataset that was not used in the model training. The performance of models is measured in terms of mean absolute error (MAE), mean bias error (MBE) and the coefficient of determination (R^2) determined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{3.4}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \tag{3.5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3.6}$$

Where N is the number of observations; y is the measured value; $\hat{y}$ is the predicted value; and $\bar{y}$ is the mean of measured values.

The MAE indicates the average deviation of predictions from the measured value with smaller values indicating better performance. The MBE measures the average systematic bias, positive or negative values indicate the average tendency of the predicted values to be larger or smaller than the measured values, respectively. The R^2 measures the correspondence between predicted and measured data with higher values indicating stronger correspondence. The MAE was chosen over the root mean squared error (RMSE) as RMSE may be an inappropriate measure for averaging (Willmott & Matsuura, 2005).

4.3.5.3.2 Variable importance

The predictor variable importance is the statistical significance of each predictor variable with respect to its effect on the generated model. For the tree-based models, RF and BRT, variable importance is calculated internally within the model algorithm (Equation 1). For the rest of the ML models, we calculated the predictor variable importance by recursive feature elimination method, which is done by recursively removing predictors before training a model and evaluating the change in model performance. In this method, to account for possible bias in variable subset selection (Ambroise & McLachlan, 2002; Hastie et al., 2009), we included a separate layer of 10-fold cross-validation to the entire sequence modeling steps.

4.3.5.3.3 Effect of predictor variables

The relationship between the predictor variables and outputs for a black-box model can be analyzed using model-independent methods such as partial dependence plots or accumulated local effects (ALE) plots (Apley, 2016; Greenwell, 2017). These plots help explain the relationship between the outcome of black-box supervised ML models and predictors of interest. We use the ALE plots to analyze the effect of selected predictor variables. Although similar, the ALE plots are preferred over partial dependence plots for their speed and their ability to produce unbiased plots when variables are correlated (Apley, 2016). The value of the ALE is centered so that the mean effect is zero; it can be interpreted as the effect of the variable on the outcome at a certain value compared to the average prediction of the data. For example, an ALE estimate of -2 when a variable of interest has value 3, then the prediction is lower by 2 compared to the average prediction (Molnar, 2019).

4.3.6 Predicting Soil Moisture

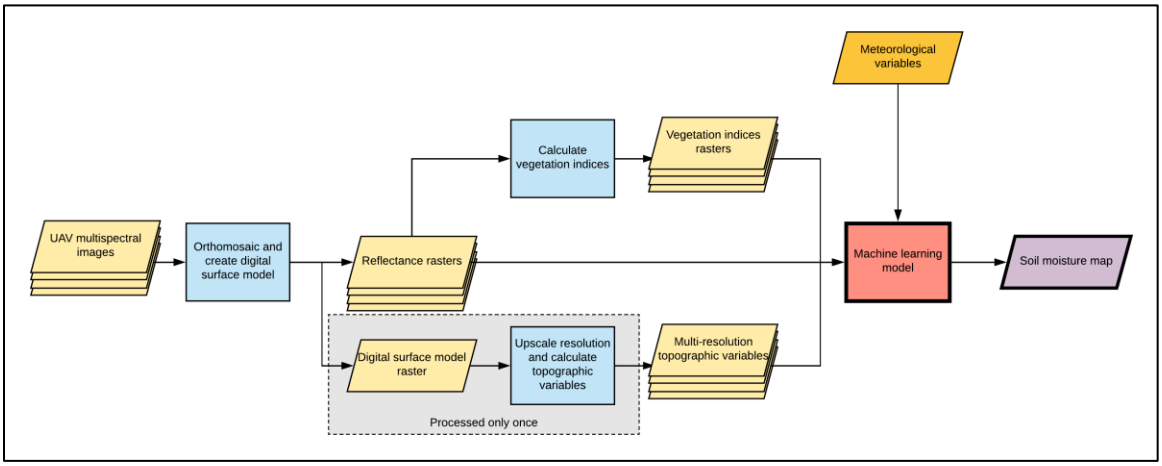

Figure 4-11 Process flow chart of predicting surface soil moisture from new UAS multispectral images.

4.3.7 Software Used

Preliminary UAS image processing— radiometric calibrations and, orthomosaic and DEM generation—is done using Pix4D photogrammetry software (Pix4D, Lausanne, Switzerland). Raster rescaling, terrain analysis, and spatial data visualization are done in ArcGIS Pro software (ESRI, Redlands, CA, USA). The machine learning process—data preparation, model tuning, and prediction—are done in R environment (R, Vienna, Australia).

4.4 Results

4.4.1 Model Performance

The performances of the five machine learning algorithms we tested, measured in total absolute residuals, are reported in RF Figure 4-12. Overall, BRT and RF algorithms better than the remaining three (RVR, SVR, and ANN). Therefore, in the remainder of this chapter, we will focus mainly on the BRT models. The BRT algorithm had the best prediction accuracy closely followed by RF. This is consistent with other studies that find ensemble decision-tree-based regression models such as RF and BRT outperform many ML algorithms (Caruana & Niculescu-Mizil, 2006); particularly in terrain and soil spatial predictions (Hengl et al., 2017, 2018; Keskin et al., 2019; Nussbaum et al., 2018; Szabó et al., 2019). The RF model is much easier and faster to train compared to the other ML algorithms used. Since the ensemble trees are independent in RF model, the 'forest' can be grown in simultaneously, which dramatically increases processing efficiency in parallel computing. In addition, the RF model has few hyperparameters to tune. In contrast, the ensemble trees in the BRT algorithm must be grown sequentially since each new tree is dependent on the previous ensemble (which makes parallel processing challenging).

Training a BRT model requires tuning multiple hyperparameters—seven in our implementation of the BRT model.

The individual performances of the 30 tuned BRT models on their respective testing sets are shown in Figure 4-13. On average the ensemble of the BRT models had MAE of 3.77 % across all 30 testing sets. Given that the locations of ground sampling points were clustered in the six transects, selection of testing and training datasets can lead to spatial bias. To minimize such biases, the transects for the testing dataset were randomly for each sampling day. Comparison of the performance of models across the different training-testing splits suggests that the bias based on testing set selection was minimal.

The measured water contents from the testing data sets are compared with the prediction by all the trained models in Figure 4-14.

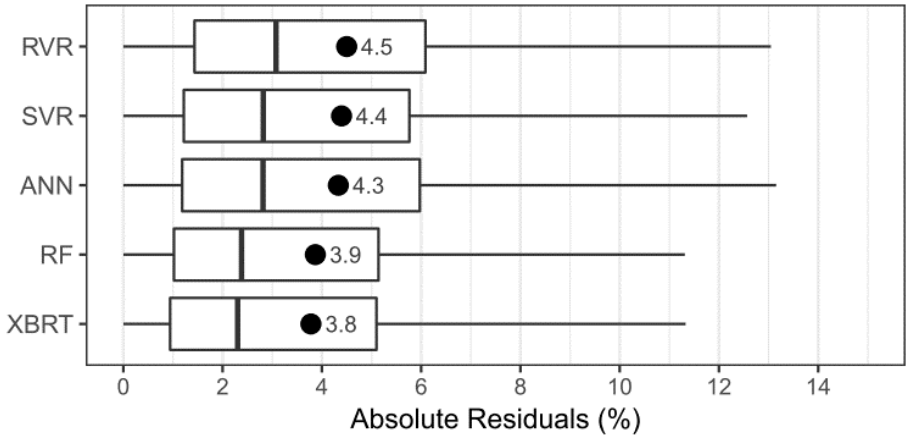

Figure 4-12 Distribution of residuals and MAE on the testing set by machine learning algorithm. Filled circles and values to their right indicate the average MAE.

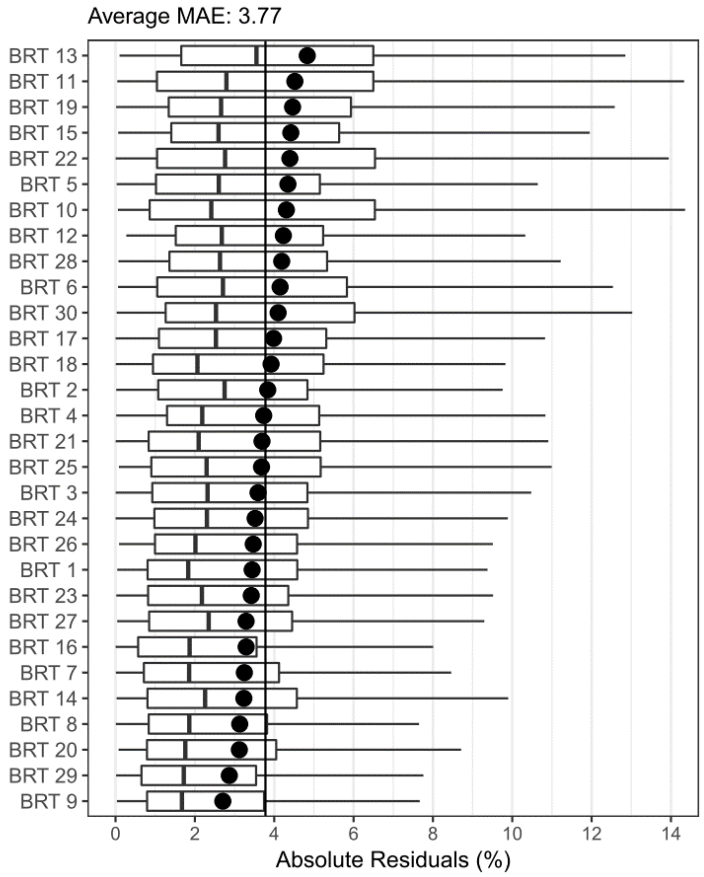

Figure 4-13 Distribution of residuals and MAE on the testing set for the 30 BRT models. Filled circles indicate MAE and solid vertical line indicates the average MAE overall individual models.

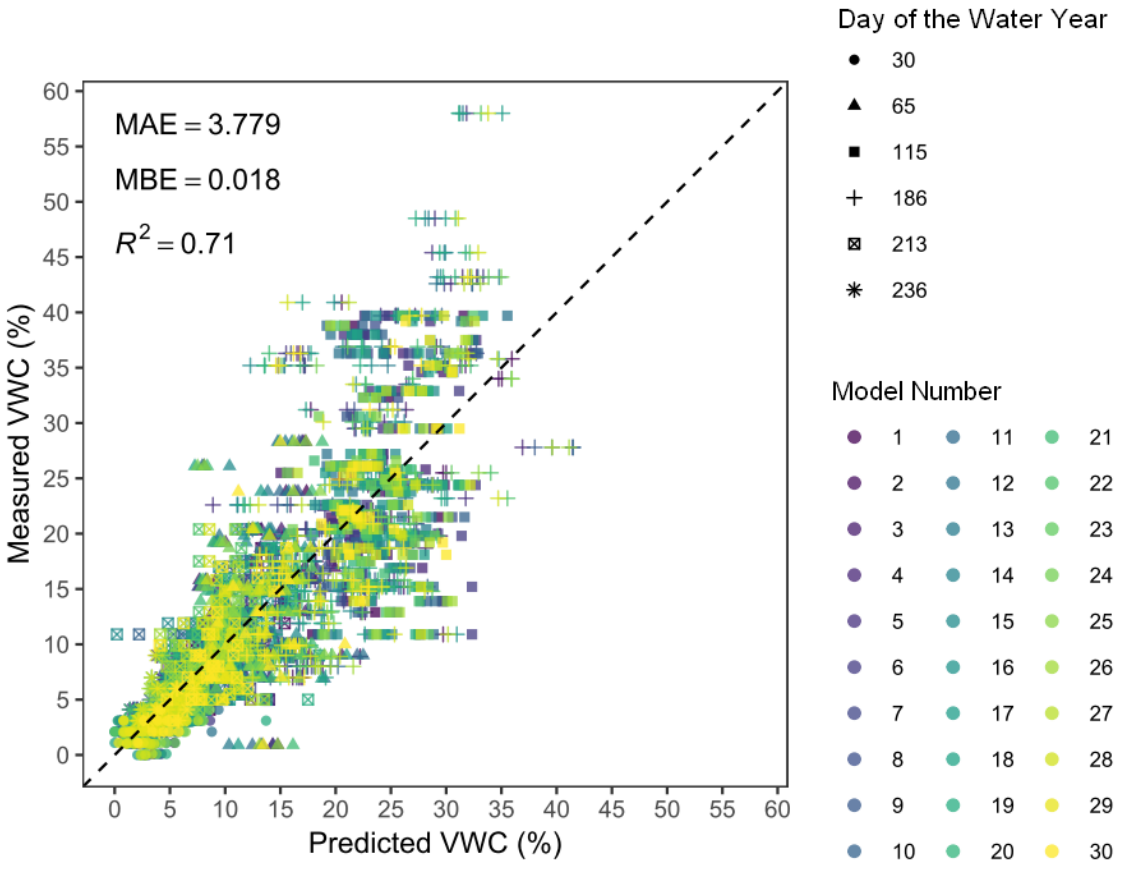

Figure 4-14 Scatter of the measured versus predicted soil water content of the testing sets around 1:1 line. MAE, MBE, and R^2 are averaged across the 30 models.

4.4.2 Predictor Variable Importance

The relative importance of predictors from the BRT model grouped by variable type (lumping variables regardless of variable specifications such as summation window for precipitation or pixel resolution for the raster) are shown in Figure 4-15. Notice that the only temporally dynamic variables are the meteoric and reflectance variables, all the topographic variables need to be generated only one time for the study area. As can be seen in Figure 4-15, the top four important variables are precipitation, reflectance in the red band, PET, and TPI. The topmost important variables not grouped by type are the 15 and 30-day cumulative precipitation, 30-day cumulative PET, and the red bands. TPI and flow accumulation are the most important of the topographic variables (Appendix C, Figure A0-12).

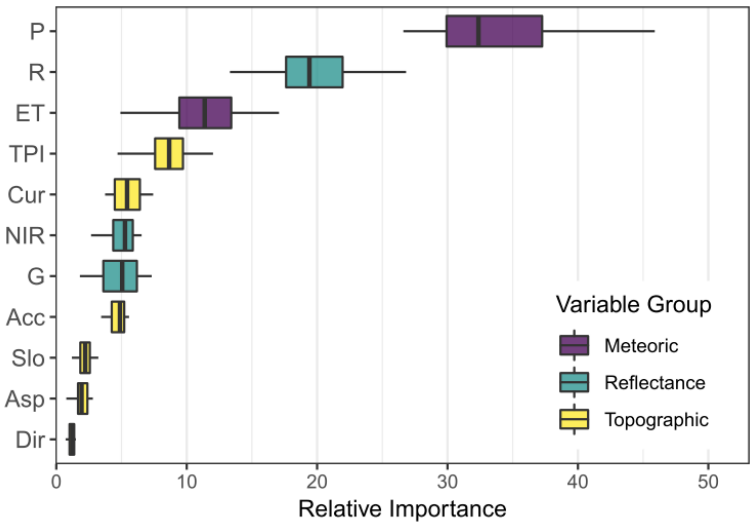

Figure 4-15 Sum of the relative variable importance distribution of the 30 models grouped by variable type.

4.4.3 Effect of Predictor Variables

We used the ALE plots to investigate the effects of TPI, curvature, and flow accumulation variables on soil moisture estimates (Figure 4-16). Given the high importance of these topographic variables, we were interested in identifying the nature of each variable's relation with soil moisture and thresholds of large changes. Predicted soil moisture generally increased with flow accumulation. At 1 m resolution DEM, soil moisture initially decreased as surface become less convex and increased as surface curvature transitioned from convex to concave (values -5 to +5). Above the value of about +5, soil moisture decreased with curvature. However, at lower scales (3, 5 and 50 m resolution DEM), soil moisture was maximum at convex to concave transition (near 0) but there was no decrease in moisture at higher concave values and with the lowest resolution (50 m DEM) soil moisture continued to increase with an increase in concavity of surface). TPI is very scale-dependent. However, TPI across all scales had pattern relation with soil moisture. Negative TPI values indicate trends towards valleys, zero values indicate flat areas if the slope is shallow or mid-slope areas for areas with significant slope and positive TPI values indicate trends towards ridgetops (Jenness et al., 2013). Across all scales, there was a negative relationship between TPI and soil moisture at negative values and a positive relationship at positive TPI values; this indicates valleys and ridge tops were wetter than mid-slope areas. The TPI scale that had was the highest variable importance was calculated with an inner diameter of 15 m and an outer diameter of 35 m (TPI (15, 35)). At this scale, the decrease in soil moisture moving from valleys to mid-slope areas was most pronounced.

Topography has a strong control on soil moisture distribution at landscape scales (Sørensen et al., 2006). While the TPI was the most important topographic variable in determining soil moisture, the TWI was found to be not an important predictor. Although the ML

models are considered non-spatial models, that is, they do not consider sampling location information and spatial autocorrelations (Georganos et al., 2019; Hengl et al., 2018). The inclusion of spatially dependent variables (specifically: curvature, flow accumulation, and TPI) as predictors, however, means that the models do account for spatial information to an extent and this should make the predictions more spatially relevant.

The red band was the most predictive of the three bands. Reflectance in the red-edge was found to be not an important predictor. The two greenness indices we tested, i.e., NDVI and TTVI, were not important even though their constituent bands, the red and NIR were important. The lower importance of NIR in perdition surface soil moisture was particularly surprising given the sensitivity of NIR band to moisture to plant stress.

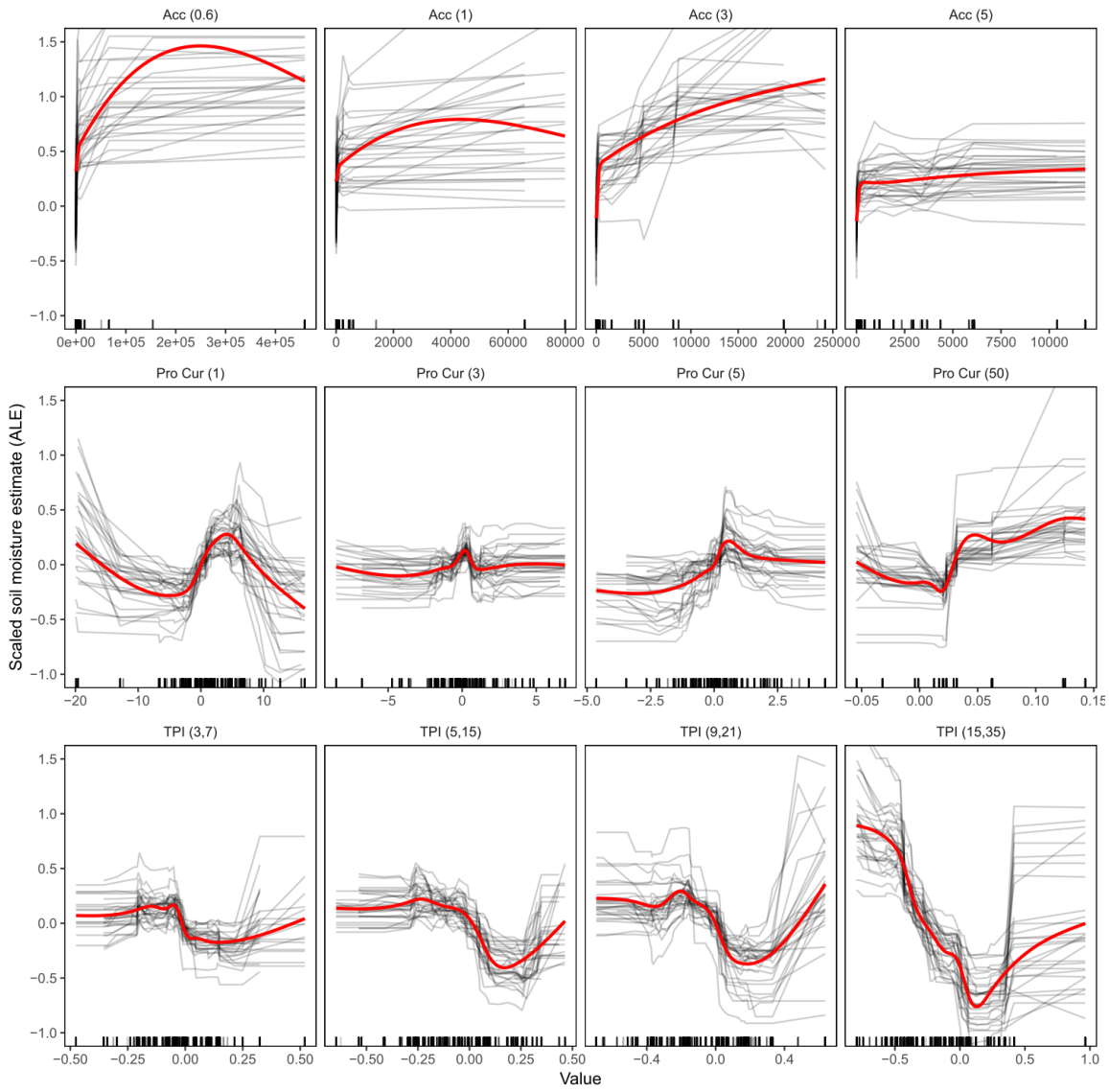

Figure 4-16 ALE plots for flow accumulation, profile curvature, and TPI variables at four different scales (see Table 4-2). Black curves represent individual effects of the 30 models, and red curves are smoothed trendlines overall individual models. Marks along the x-axis show the distribution of data in the model training set.

4.4.4 Spatial Prediction of Soil Moisture

Volumetric water content from the best BRT model was used to predict soil moisture for the test area for all six measurement days Figure 4-17. While the mean moisture content was mostly determined by the day, the distribution shows a visually similar distribution to topography. Ridges appear drier while valleys appear wetter.

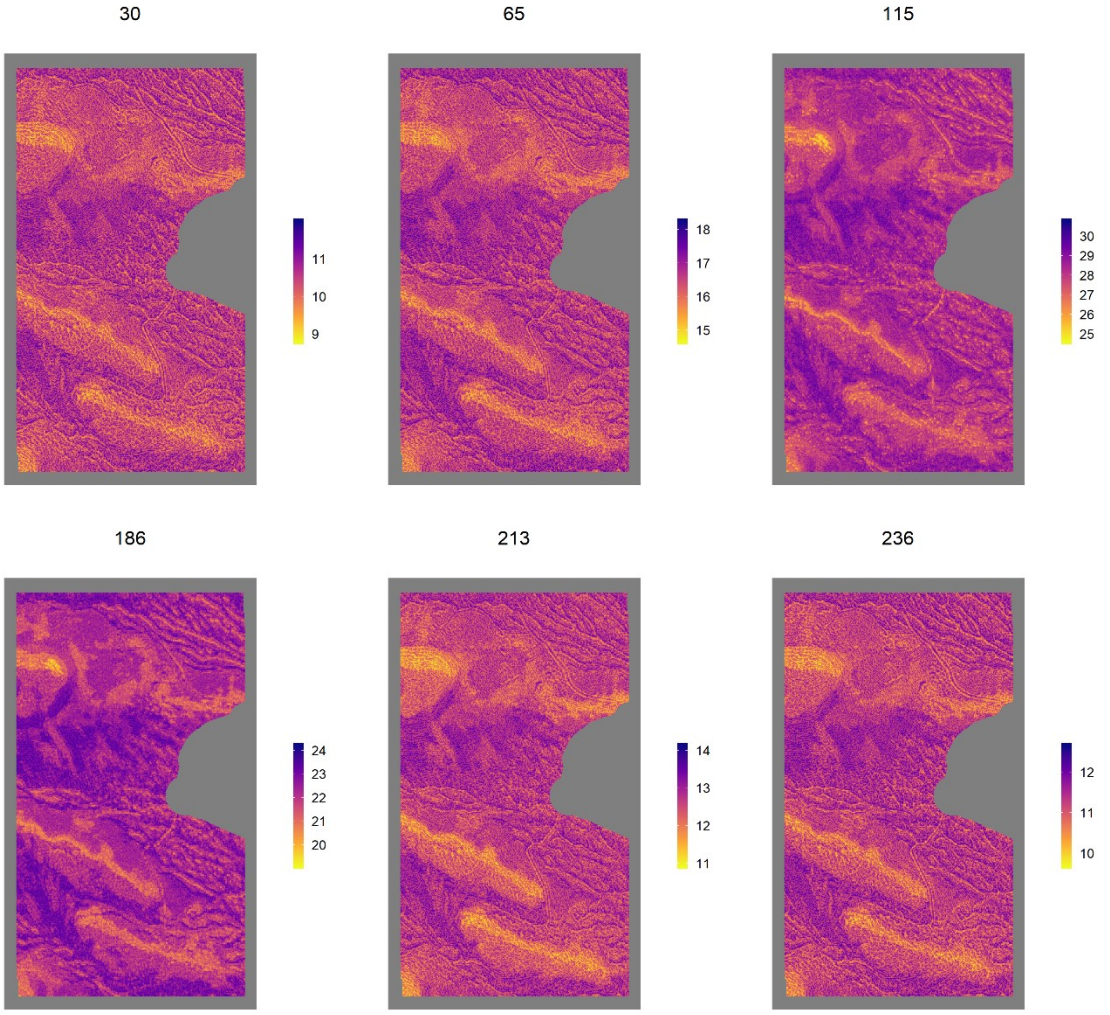

Figure 4-17 Predicted volumetric water content (%) over the study area for the six days sampled. Days of the water year 30, 65, 115, 186, 213, and 236 are 10/30/2017, 12/4/2017, 1/23/2018, 4/4/2018, 5/1/2018, and 5/24 /2018, respectively).

The distribution of soil moisture predictions over the test area for each day is shown in Figure 4-18. Soil moisture shows multimodal distribution except for the wettest water year day (115).

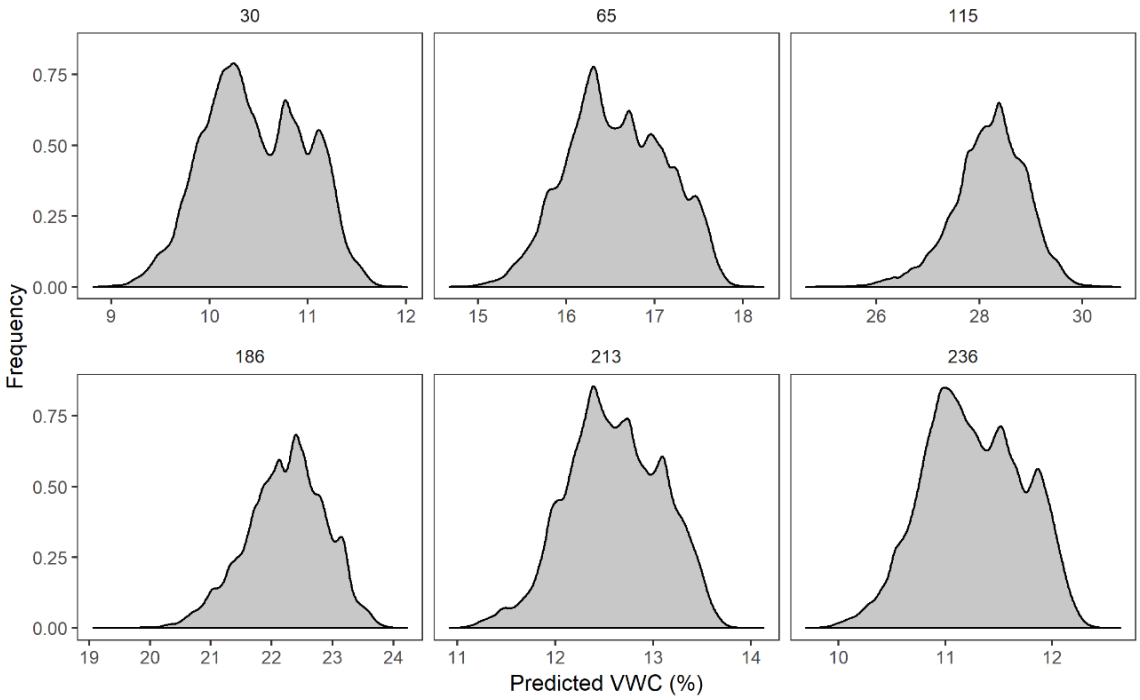

Figure 4-18 Kernel density estimate of predicted soil moisture over the study area for the six sampling days.

A close-up map of soil moisture prediction for January 23, 2018, is shown in Figure 4-19 and shows that soil moisture varies considerably with topography.

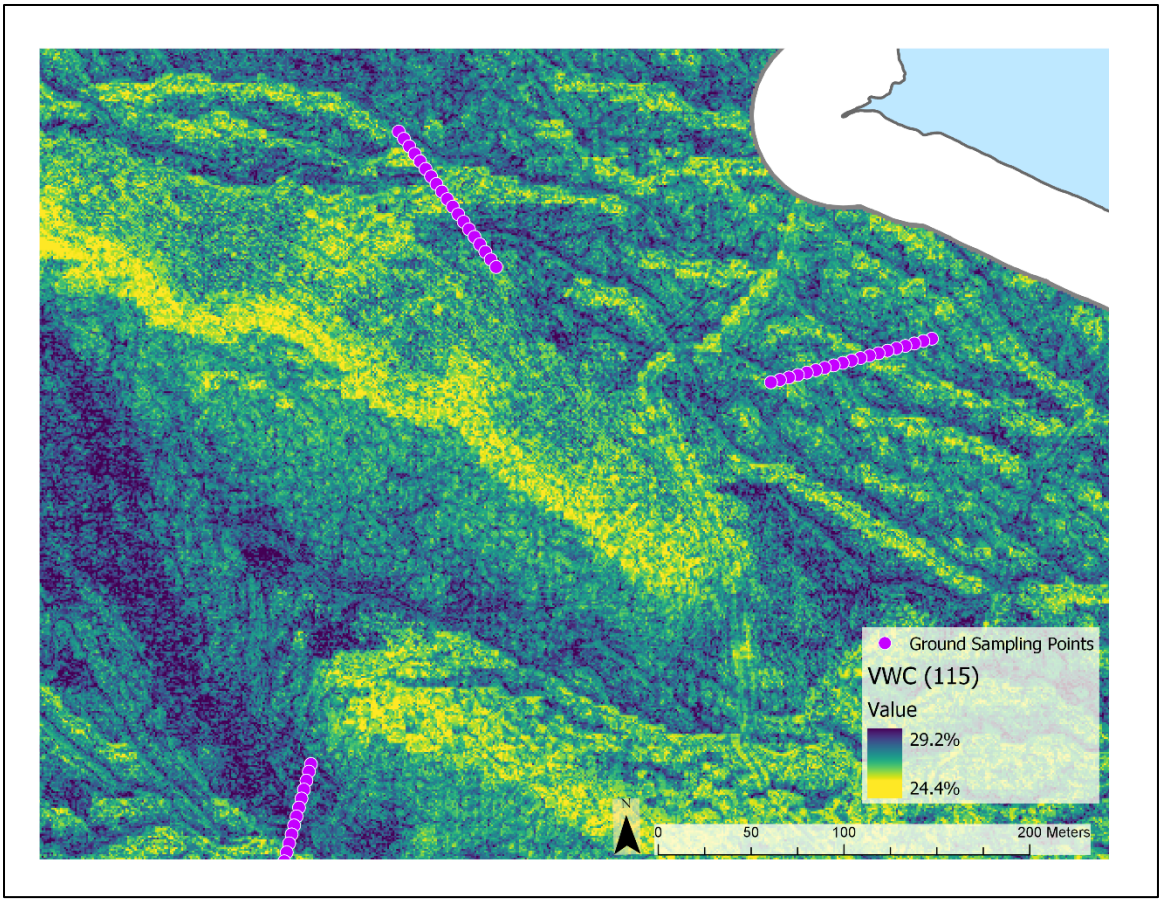

Figure 4-19 Predicted volumetric soil water content (%) map for 115th day of the water year (January 23, 2018).

4.5 Conclusion and Outlook

This research serves as a proof of concept: surface soil moisture can be interpreted with high accuracy from multispectral UAS remote sensing using machine learning methods. As a data mining technique, machine learning model performance and reliability are closely tied to the quantity of data. Although the number and spatial coverage of ground sampling points were reasonable, the addition of more sampling points and some occasional measurements at random locations within the study area would have greatly helped to strengthen the reliability of the models. Multi-year studies are eventually needed to ensure that the model can be used for future predictions. It will be interesting to see the importance of precipitation variables versus reflectance on models developed on multiple year data. The reflectance variables, for instance, maybe more dependent on the annual cycle of vegetation than meteoric variables, and it would be interesting to see how variable importance might change. Although more challenging, studies of deeper soil root-zone soil moisture are more ecologically relevant. Similar studies with root-zone soil moisture should be done. Expanding the reflectance information beyond multispectral bands could lead to important improvement in soil moisture prediction. Although lightweight thermal or hyperspectral sensors are currently very expensive and may not be financially feasible

for routine application at this time, it is possible that those technologies might become more affordable in the future.

The possibility of using the high-resolution topographic variables from UAS with reflectance data from satellite remote sensing is an interesting topic to study. This would mean the UAS would only be flown once in an area and future and past predictions of soil moisture can be estimated using satellite images and meteoric data from stations. This would also be ideal to downscale satellite data by integrating high-resolution topographic information from UAS remote sensing. Although the ML models are high performing and generalizable models, they are non-spatial models that do not consider sampling location information and spatial autocorrelations. This can potentially compromise the model ability to appropriately address spatial heterogeneity (Georganos et al., 2019; Hengl et al., 2018). Hengl et al. (2018) introduce a novel method to incorporate spatial information into a non-spatial ML model by including distances between sampling points as predictor variables, and they show this method (although still in its formative stage) has comparable accuracy to kriging methods. The potential to improve soil moisture predictions by using such spatially integrated methods should be considered in future research.

Acknowledgments

I am grateful to Anna Fryjoff-Hung and Andreas Anderson (UC Merced) for their expertise and collaboration which enabled me to conduct this study. This work was made possible by the support of U.S. Fish and Wildlife Service Agreement #P1740401 as administered by the California Department of Fish and Wildlife; UC Natural Reserve System; UC Merced Physical and Environmental Planning; and UC Merced Vernal Pools and Grassland Reserve. We gratefully acknowledge computing time on the Multi-Environment Computer for Exploration and Discovery (MERCED) cluster at UC Merced, which was funded by National Science Foundation Grant No. ACI-1429783.

References

Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources*, *33*(1), 69–80. https://doi.org/10.1016/j.advwatres.2009.10.008

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., & Notarnicola, C. (2015). Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sensing*, *7*(12), 16398–16421. https://doi.org/10.3390/rs71215841

Ambroise, C., & McLachlan, G. J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences*, *99*(10), 6562–6566. https://doi.org/10.1073/pnas.102102699

Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, *11*(3), 138–146. https://doi.org/10.1890/120150

Apley, D. W. (2016). *Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models*. *arXiv preprint*. Retrieved from https://arxiv.org/ftp/arxiv/papers/1612/1612.08468.pdf

Barrett, B. W., & Petropoulos, G. P. (2014). Satellite Remote Sensing of Surface Soil Moisture. In G. P. Petropoulos (Ed.), *Remote Sensing of Energy Fluxes and Soil Moisture Content* (pp. 85–120). CRC Press. https://doi.org/doi:10.1201/b15610-6

Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., & Sommer, S. (2009). Using Imaging Spectroscopy to study soil properties. *Remote Sensing of Environment*, *113*(SUPPL. 1), S38–S55. https://doi.org/10.1016/j.rse.2008.09.019

Ben-Shimon, D., & Shmilovici, A. (2006). Kernels for the Relevance Vector Machine - An Empirical Study. In M. Last, P. S. Szczepaniak, Z. Volkovich, & A. Kandel (Eds.), *Advances in Web Intelligence and Data Mining* (pp. 253–263). Springer-Verlag GmbH. https://doi.org/10.1007/3-540-33880-2

Berni, J. A. J., Zarco-Tejada, P. J., Suárez, L., González-Dugo, V., & Fereres, E. (2009). Remote sensing of vegetation from UAV platforms using lightweight multispectral and thermal imaging sensors. *Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci*, *38*, 6 pp. https://doi.org/10.1007/s11032-006-9022-5

Breiman, L. (2001). Random Forest. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

California Department of Water Resources. (2018). UC Merced Weather Station. Retrieved February 1, 2019, from http://cdec.water.ca.gov/dynamicapp/staMeta?station_id=UCM

California Irrigation Management Information System. (n.d.). Data Overview. Retrieved April 19, 2019, from https://cimis.water.ca.gov/Resources.aspx

California Irrigation Management Information System. (2018). CIMIS Station Report. Retrieved February 1, 2019, from https://cimis.water.ca.gov/WSNReportCriteria.aspx

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning - ICML '06* (pp. 161–168). New York, New York, USA: ACM Press. https://doi.org/10.1145/1143844.1143865

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16* (pp. 785–794). New York, New York, USA: ACM Press. https://doi.org/10.1145/2939672.2939785

Chen, X., & Jeong, J. C. (2007). Enhanced Recursive Feature Elimination. *Proceedings - 6th International Conference on Machine Learning and Applications, ICMLA 2007*, 330–335. https://doi.org/10.1109/ICMLA.2007.35

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *92*, 79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013

Cortes, C., & Vapnik, V. N. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Das, N. N., & Mohanty, B. P. (2006). Root Zone Soil Moisture Assessment Using Remote Sensing and Vadose Zone Modeling. *Vadose Zone Journal*, *5*(1), 296. https://doi.org/10.2136/vzj2005.0033

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. N. (1997). Support vector regression machines. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9* (Vol. 1, pp. 155–161). MIT Press. Retrieved from http://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf

Elarab, M. (2016). *The Application of Unmanned Aerial Vehicle to Precision Agriculture: Chlorophyll, Nitrogen, and Evapotranspiration Estimation*. Utah State University.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1017/CBO9781107415324.004

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Fryjoff-Hung, A. F. (2018). *2D Hydrodynamic Modeling for Evaluating Restoration Potential of a Vernal Pool Complex*. University of California, Merced.

Georganos, S., Grippa, T., Gadiaga, A. N., Linard, C., Lennert, M., Vanhuysse, S., … Kalogirou, S. (2019). Geographical Random Forests: A Spatial Extension of the Random Forest Algorithm to Address Spatial Heterogeneity in Remote Sensing and Population Modelling. *Geocarto International*, 1–12. https://doi.org/10.1080/10106049.2019.1595177

Greenwell, B. M. (2017). pdp: An R Package for Constructing Partial Dependence Plots. *The R Journal*, *9*(1). Retrieved from https://github.com/bgreenwell/pdp/issues.

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, *3*(3), 1157–1182.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 389–422. https://doi.org/10.1023/A:1012487302797

Hassan-Esfahani, L., Torres-Rua, A., Jensen, A., & McKee, M. (2015). Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. *Remote Sensing*, *7*(3), 2627–2646. https://doi.org/10.3390/rs70302627

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). New York, NY: Springer New York. https://doi.org/10.1007/b94608

Haubrock, S.-N., Chabrillat, S., Lemmnitz, C., & Kaufmann, H. (2008). Surface soil moisture quantification models from reflectance data under field conditions. *International Journal of Remote Sensing*, *29*(1), 3–29. https://doi.org/10.1080/01431160701294695

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., … Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. https://doi.org/10.7717/peerj.5518

Hillel, D. (1998). *Environmental Soil Physics*. San Diego, CA: Academic Press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Jenness, J., Brost, B., & Beier, P. (2013). Land Facet Corridor Designer. Retrieved from www.corridordesign.org

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab: An S4 Package for Kernel Methods in R (version 0.9-25). *Journal of Statistical Software*, *11*(9), 1–20. Retrieved from http://www.jstatsoft.org/v11/i09/

Keskin, H., Grunwald, S., & Harris, W. G. (2019). Digital mapping of soil carbon

fractions with machine learning. *Geoderma*, *339*, 40–58. https://doi.org/10.1016/j.geoderma.2018.12.037

Kuhn, M. (2017). The caret Package. Retrieved from http://topepo.github.io/caret/index.html

Liaw, A., & Wiener, M. (2015). Package "randomForest" (v.4.6-12). *R Package*. https://doi.org/10.5244/C.22.54

Malley, D. F., Martin, P. D., & Ben-Dor, E. (2004). Application in Analysis of Soils. In R. Craig, R. Windham, & J. Workman (Eds.), *Near-Infrared Spectroscopy in Agriculture* (pp. 729–784). Madison, WI, USA: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Matei, O., Rusu, T., Petrovan, A., & Mihuţ, G. (2017). A Data Mining System for Real Time Soil Moisture Prediction. *Procedia Engineering*, *181*, 837–844. https://doi.org/10.1016/j.proeng.2017.02.475

Molnar, C. (2019). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Retrieved from https://christophm.github.io/interpretable-ml-book/

Muller, E., & Décamps, H. (2000). Modeling soil moisture–reflectance. *Remote Sensing of Environment*, *76*(2), 173–180. https://doi.org/10.1016/S0034-4257(00)00198-X

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*(DEC), 1–11. https://doi.org/10.3389/fnbot.2013.00021

Nichols, S., Zhang, Y., & Ahmad, A. (2011). Review and evaluation of remote sensing methods for soil-moisture estimation. *Journal of Photonics for Energy*, *2*(1), 028001. https://doi.org/10.1117/1.3534910

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., … Papritz, A. (2018). Evaluation of digital soil mapping approaches with large sets of environmental covariates. *SOIL*, *4*(1), 1–22. https://doi.org/10.5194/soil-4-1-2018

Ochsner, T. E., Cosh, M. H., Cuenca, R. H., Dorigo, W. a., Draper, C. S., Hagimoto, Y., … Zreda, M. (2013). State of the Art in Large-Scale Soil Moisture Monitoring. *Soil Science Society of America Journal*, *77*(6), 1888. https://doi.org/10.2136/sssaj2013.03.0093

Pachepsky, Y. A., Timlin, D., & Varallyay, G. (1996). Artificial Neural Networks to Estimate Soil Water Retention from Easily Measurable Data. *Soil Science Society of America Journal*, *60*(3), 727–733. https://doi.org/10.2136/sssaj1996.03615995006000030007x

Paloscia, S., Pampaloni, P., Pettinato, S., & Santi, E. (2008). A Comparison of Algorithms for Retrieving Soil Moisture from ENVISAT/ASAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(10), 3274–3284. https://doi.org/10.1109/TGRS.2008.920370

Petropoulos, G. P., Ireland, G., & Barrett, B. W. (2015). Surface soil moisture retrievals

from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth*, *83–84*, 36–56. https://doi.org/10.1016/j.pce.2015.02.009

Pix4D. (n.d.). Learn more about Sequoia. Retrieved November 14, 2016, from https://pix4d.com/sequoia-faq/

Price, J. C. (1977). Thermal inertia mapping: A new view of the earth. *Journal of Geophysical Research*, *82*, 2582–2590.

Price, J. C. (1985). On the analysis of thermal infrared imagery: The limited utility of apparent thermal inertia. *Remote Sensing of Environment*, *18*(1), 59–73. https://doi.org/10.1016/0034-4257(85)90038-0

Rana, G., & Katerji, N. (2000). Measurement and estimation of actual evapotranspiration in the field under Mediterranean climate: a review. *European Journal of Agronomy*, *13*(2–3), 125–153. https://doi.org/10.1016/S1161-0301(00)00070-8

Ridgeway, G. (2012). *Generalized Boosted Models: A guide to the gbm package*. *Compute*. https://doi.org/10.1111/j.1467-9752.1996.tb00390.x

Schaap, M. G., Leij, F. J., & van Genuchten, M. T. (2001). Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, *251*(3–4), 163–176. https://doi.org/10.1016/S0022-1694(01)00466-8

Schanda, E. (1986). *Physical Fundamentals of Remote Sensing*. Berlin, Germany: Springer-Verlag.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., … Teuling, A. J. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Singh, K. K., & Frazier, A. E. (2018). A meta-analysis and review of unmanned aircraft system (UAS) imagery for terrestrial applications. *International Journal of Remote Sensing*, *39*(15–16), 5078–5098. https://doi.org/10.1080/01431161.2017.1420941

Sobrino, J., Mattar, C., Jiménez-Muñoz, J. C., Franch, B., & Corbari, C. (2014). On the Synergy between Optical and TIR Observations for the Retrieval of Soil Moisture Content. In G. P. Petropoulos (Ed.), *Remote Sensing of Energy Fluxes and Soil Moisture Content* (pp. 363–390). CRC Press. https://doi.org/doi:10.1201/b15610-19

Sørensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: Evaluation of different methods based on field observations. *Hydrology and Earth System Sciences*, *10*(1), 101–112. https://doi.org/10.5194/hess-10-101-2006

Spectrum Technologies Inc. (2009). FieldScout TDR 300 Soil Moisture Meter. Plainfield, IL, USA: Spectrum Technologies, Inc.

Stark, B., McGee, M., & Chen, Y. (2015). Short wave infrared (SWIR) imaging systems using small Unmanned Aerial Systems (sUAS). In *2015 International Conference*

on Unmanned Aircraft Systems (ICUAS) (pp. 495–501). IEEE. https://doi.org/10.1109/ICUAS.2015.7152328

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., & Pásztor, L. (2019). Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. *Hydrology and Earth System Sciences*, *23*(6), 2615–2635. https://doi.org/10.5194/hess-23-2615-2019

Tian, J., Su, H., He, H., & Sun, X. (2015). An Empirical Method of Estimating Soil Thermal Inertia. *Advances in Meteorology*, *2015*, 1–9. https://doi.org/10.1155/2015/428525

Tipping, M. E. (2000). The Relevance Vector Machine. In S. A. Solla, T. K. Leen, & K. Muller (Eds.), *Advances in Neural Information Processing Systems 12* (pp. 652–658). MIT Press. Retrieved from http://papers.nips.cc/paper/1719-the-relevance-vector-machine.pdf

Torres-Rua, A., Ticlavilca, A., Bachour, R., & McKee, M. (2016). Estimation of Surface Soil Moisture in Irrigated Lands by Assimilation of Landsat Vegetation Indices, Surface Energy Balance Products, and Relevance Vector Machines. *Water*, *8*(4), 167. https://doi.org/10.3390/w8040167

Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's Global Energy Budget. *Bulletin of the American Meteorological Society*, *90*(3), 311–323. https://doi.org/10.1175/2008BAMS2634.1

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z., & Bing, Z. (2002). Relating soil surface moisture to reflectance. *Remote Sensing of Environment*, *81*(2–3), 238–246. https://doi.org/10.1016/S0034-4257(01)00347-9

Weston, J., Elisseeff, A., Schölkopf, B., & Tipping, M. (2003). Use of the Zero-Norm with Linear Models and Kernel Methods. *Journal of Machine Learning Research*, *3*, 1439–1461.

Whiting, M. L., Li, L., & Ustin, S. L. (2004). Predicting water content using Gaussian model on soil spectra. *Remote Sensing of Environment*, *89*(4), 535–552. https://doi.org/10.1016/j.rse.2003.11.009

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79–82. https://doi.org/10.3354/cr030079

Wilson, J. P., & Gallant, J. C. (2000). Digital Terrain Analysis. In J. P. Wilson & J. C. Gallant (Eds.), *Terrain Analysis: principles and applications* (pp. 1–21). https://doi.org/10.5194/hess-4-225-2000

Wong, K. (2014, January 22). Merced Vernal Pools Joins Natural Reserve System. *University of California News*. Retrieved from http://universityofcalifornia.edu/news/merced-vernal-pools-join-natural-reserve-system

Zaman, B., & Mckee, M. (2014). Spatio-Temporal Prediction of Root Zone Soil Moisture

Using Multivariate Relevance Vector Machines. *Open Journal of Modern Hydrology*, *4*(3), 80–90. https://doi.org/dx.doi.org/10.4236/ojmh.2014.43007

Zaman, B., McKee, M., & Neale, C. M. U. (2012). Fusion of remotely sensed data for soil moisture estimation using relevance vector and support vector machines. *International Journal of Remote Sensing*, *33*(20), 6516–6552. https://doi.org/10.1080/01431161.2012.690540

Zhang, D., & Zhou, G. (2016). Estimation of Soil Moisture from Optical and Thermal Remote Sensing: A Review. *Sensors*, *16*(8), 1308. https://doi.org/10.3390/s16081308

Zhang, Y., & Schaap, M. G. (2017). Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*, *547*, 39–53. https://doi.org/10.1016/j.jhydrol.2017.01.004

Zhang, Y., Schaap, M. G., & Zha, Y. (2018). A High-Resolution Global Map of Soil Hydraulic Properties Produced by a Hierarchical Parameterization of a Physically Based Water Retention Model. *Water Resources Research*, *54*(12), 9774–9790. https://doi.org/10.1029/2018WR023539

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., & Rosolem, R. (2012). COSMOS: the COsmic-ray Soil Moisture Observing System. *Hydrology and Earth System Sciences*, *16*(11), 4079–4099. https://doi.org/10.5194/hess-16-4079-2012

Chapter 5 Summary Conclusions

Through modeling, fusion of data from different sources and scales, and machine learning methods, this dissertation investigated the relationship of soil hydraulic properties and processes with the physical drivers of these attributes.

Summaries of the main sections of this dissertation are briefly given below.

Chapter 2

High-accuracy machine learning-based saturated hydraulic conductivity pedotransfer models were developed and used to describe structural perturbations on saturated hydraulic conductivity. While an increase in soil bulk density showed a general decrease in the saturated hydraulic conductivity, changes in organic carbon concentration were found to depend on soil texture. Increase in organic carbon content showed a direct relation with saturated hydraulic conductivity of fine- and medium-textured soils and an inverse relation with coarser textured soils.

Chapter 3

The long-term effect of reduced tillage and cover crop management on soil structure of a single soil was determined. Changes in pore size distribution indicated that reduced disturbance tillage with cover crop increased the proportion of smaller and larger pores while maintaining total porosity, which is consistent with an increased aggregation. The implications of management induced changes in soil structure and hydraulic properties on soil water storage were investigated using numerical irrigation simulation. Results showed that reduced tillage with cover crop increased water storage compared to standard tillage fields. While reduced tillage without cover crop showed water storage that is intermediate between reduced tillage with cover crop and standard tillage management.

Chapter 4

The control of landscape-scale surface variables on surface soil water distribution were studied by machine learning interpretation of multispectral imagery and digital elevation model collected by unmanned aircraft system platform. Machine learning models were able to interpret surface soil moisture heterogeneity reasonable accuracy. Analysis of variable importance revealed that reflectance in the red band and topographic position indices to be the two most important predictors outside of weather variables.

Appendices

Appendix A: Supplemental Figure to Chapter 2

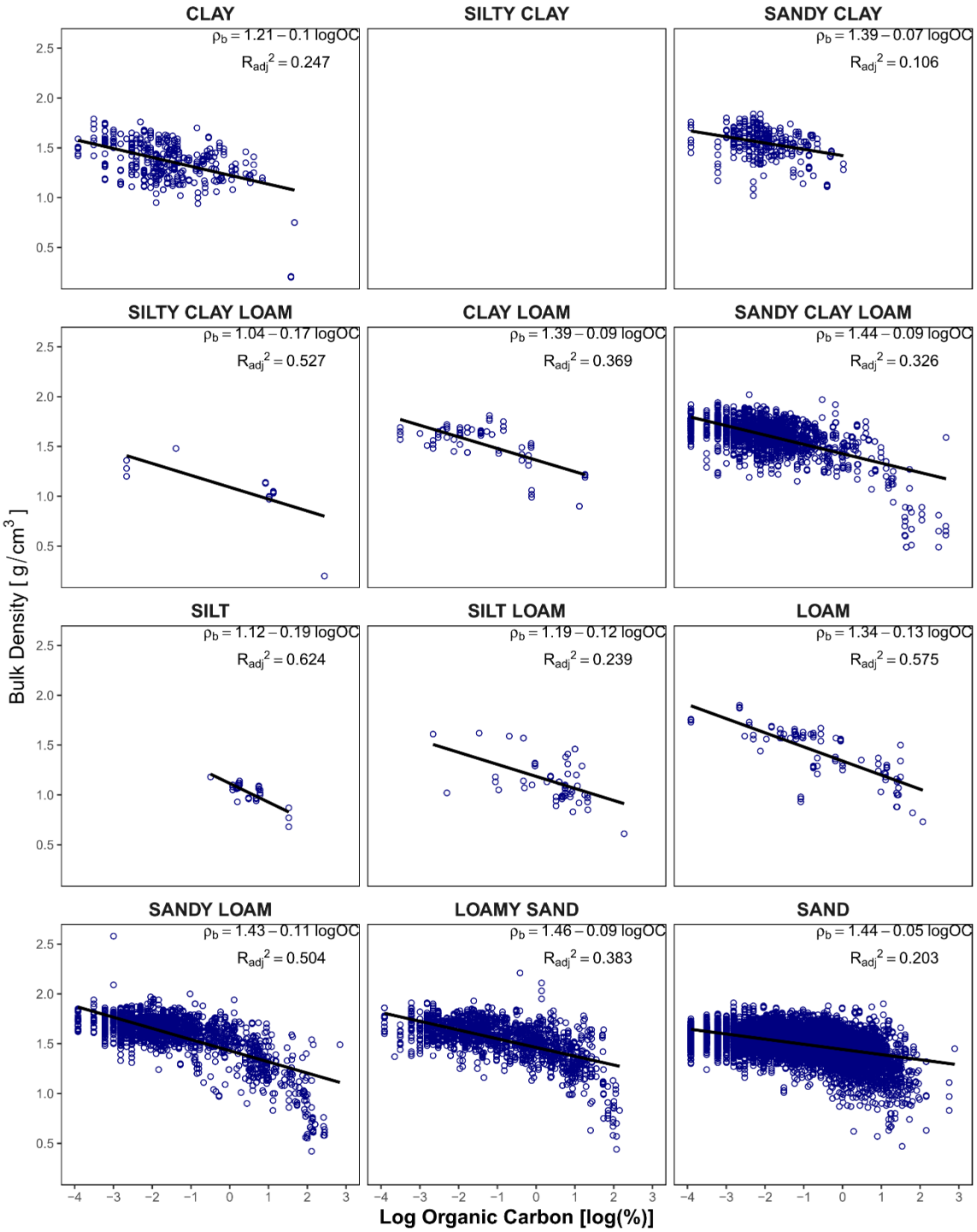

Figure A0-1 Linear correlation by soil texture class between bulk density and log-transformed organic carbon percent for the cleaned USKSAT database.

Appendix B: Supplemental Figures and Tables to Chapter 3

Supplemental Figures

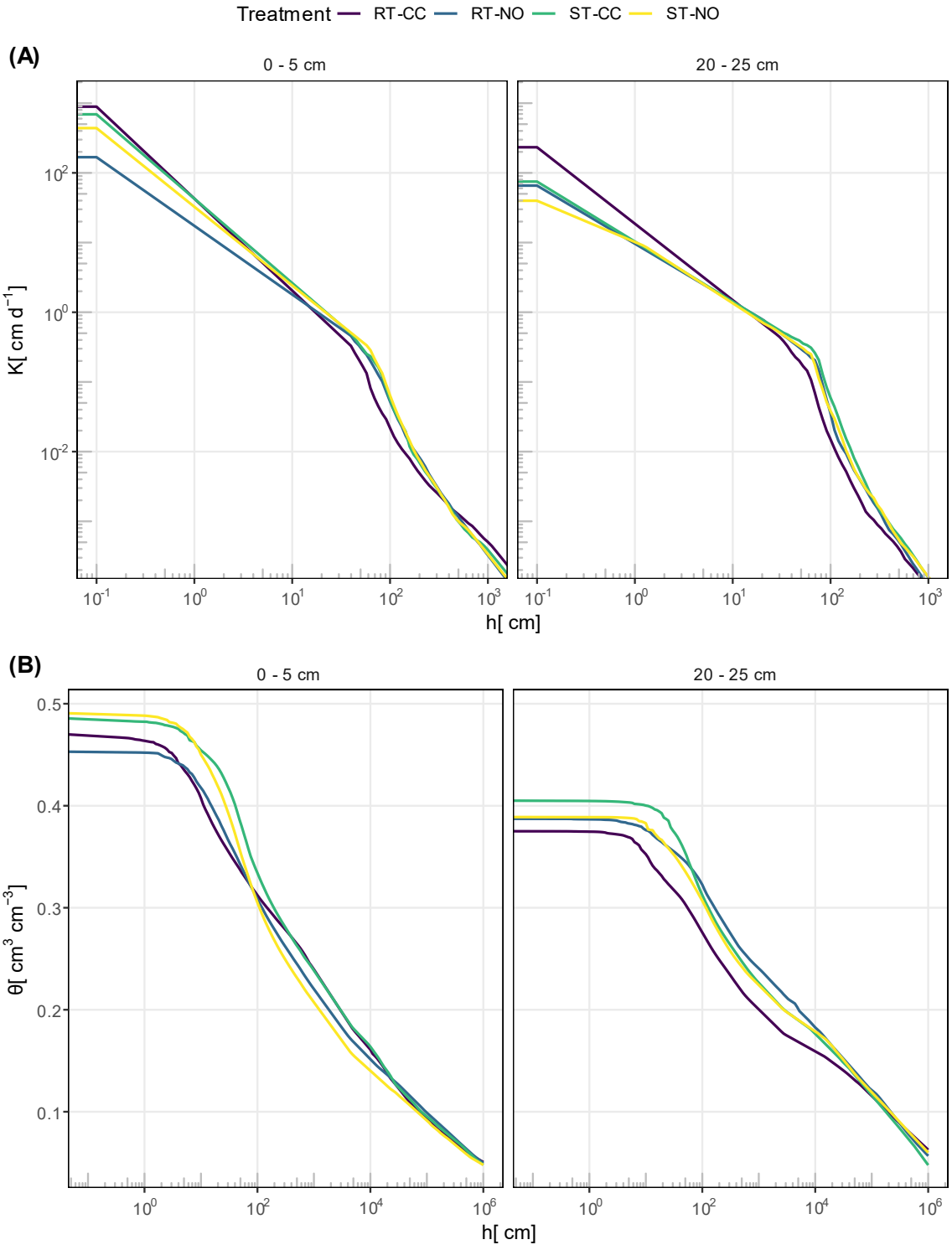

Figure A0-2 Treatment means of (A) hydraulic conductivity functions and (B) water retention curves for the top (0 – 5 cm) and subsurface (20 – 25 cm) layers.

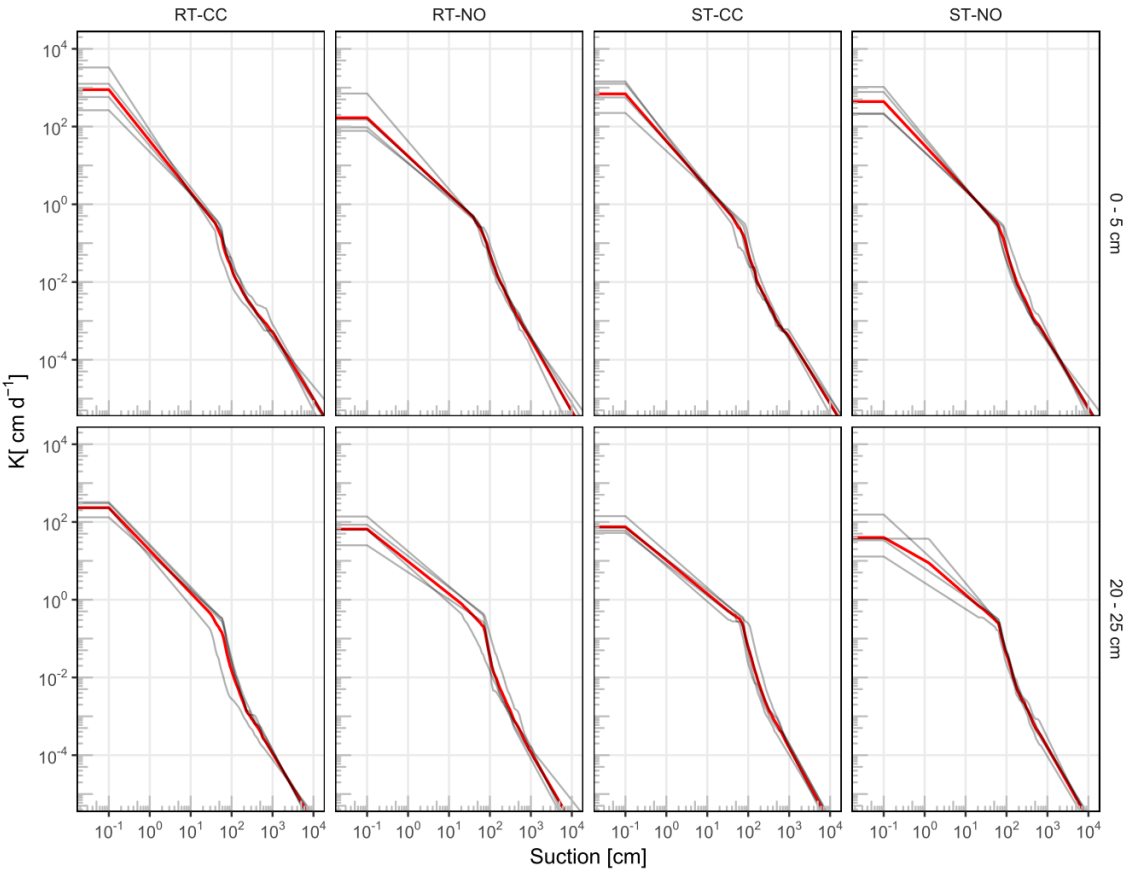

Figure A0-3 Hydraulic conductivity functions of top and subsurface layers by treatment. Grey curves are individual soil core measurements and thick red curves are means of the replicates.

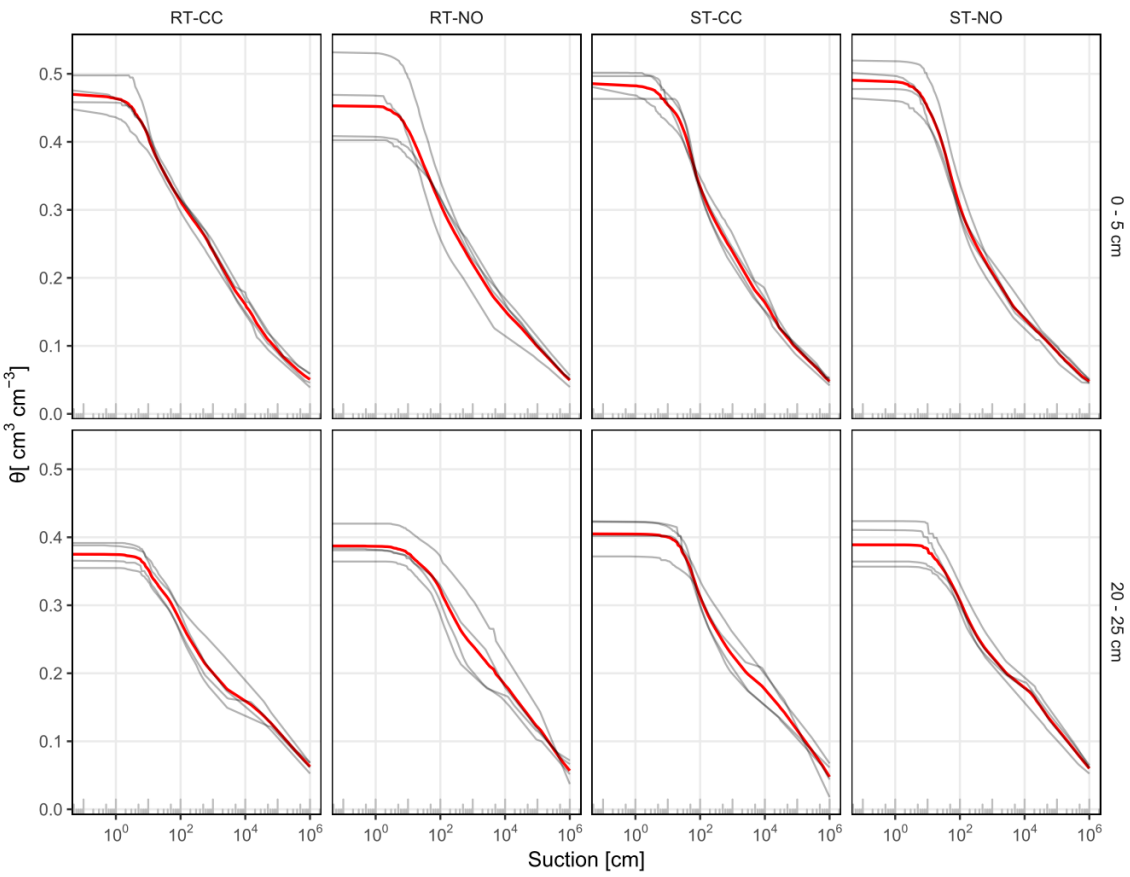

Figure A0-4Water retention curves of top and subsurface layers by treatment. Grey curves are individual soil core measurements and thick red curves are means of the replicates.

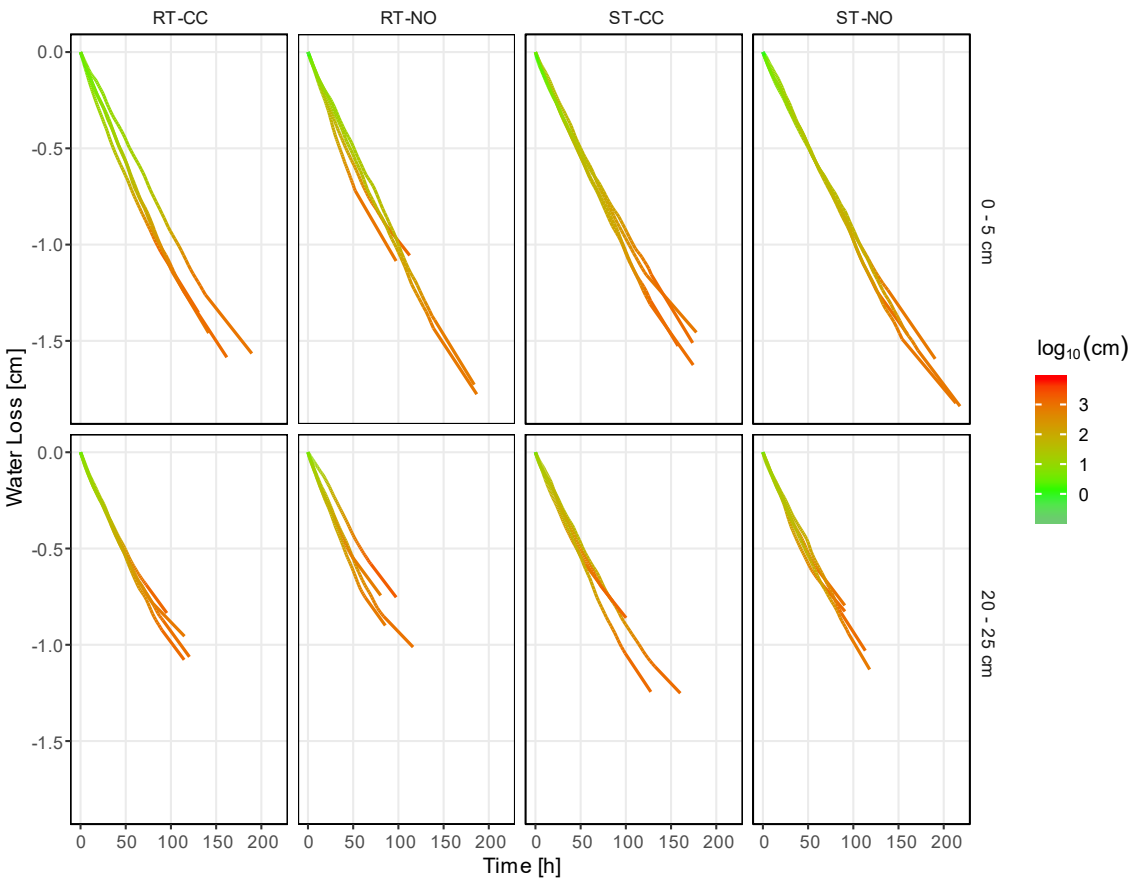

Figure A0-5 Amount of water loss over time during evaporative drying of cores. Color gradient represents suction measured 3.75 cm below soil surface.

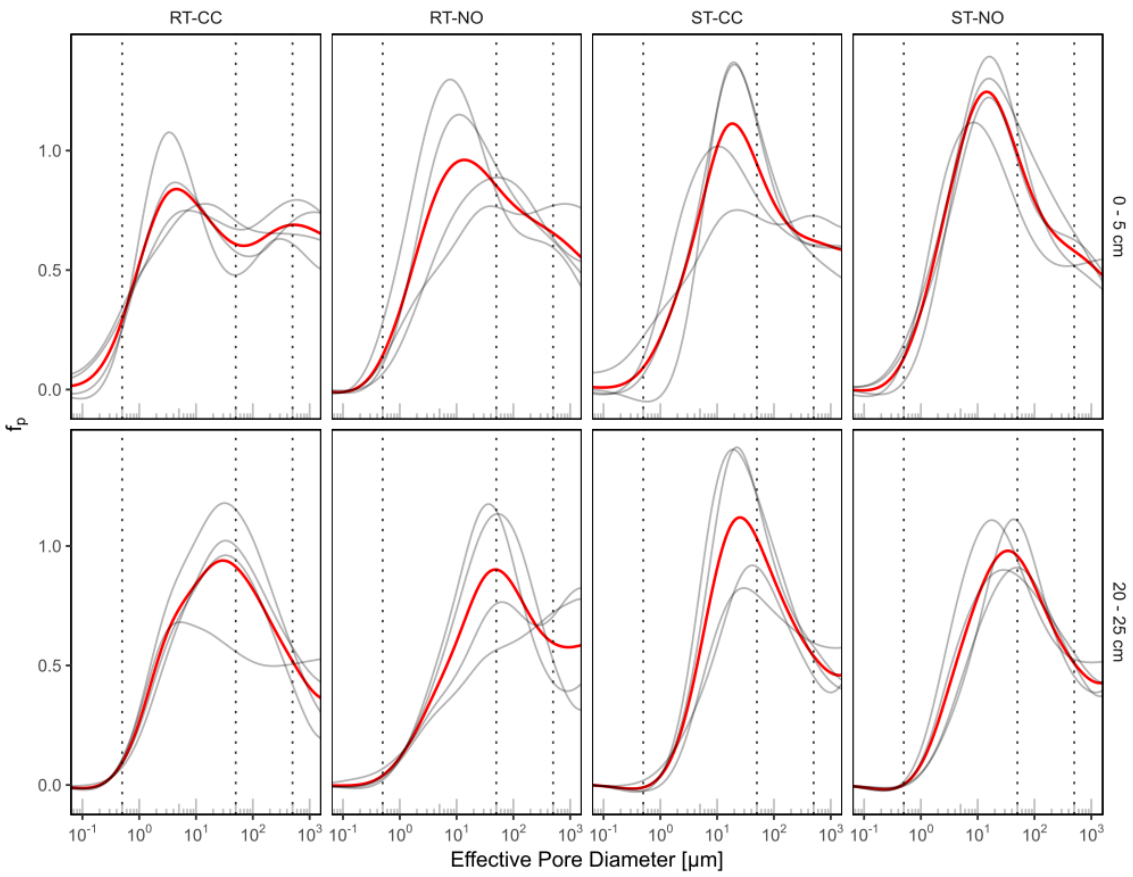

Figure A0-6 Effective pore size distribution. Grey curves are individual soil core measurements and thick red curves are means of the replicates. Vertical dotted lines indicate pore diameter sizes of 0.5, 50 and 500 µm.

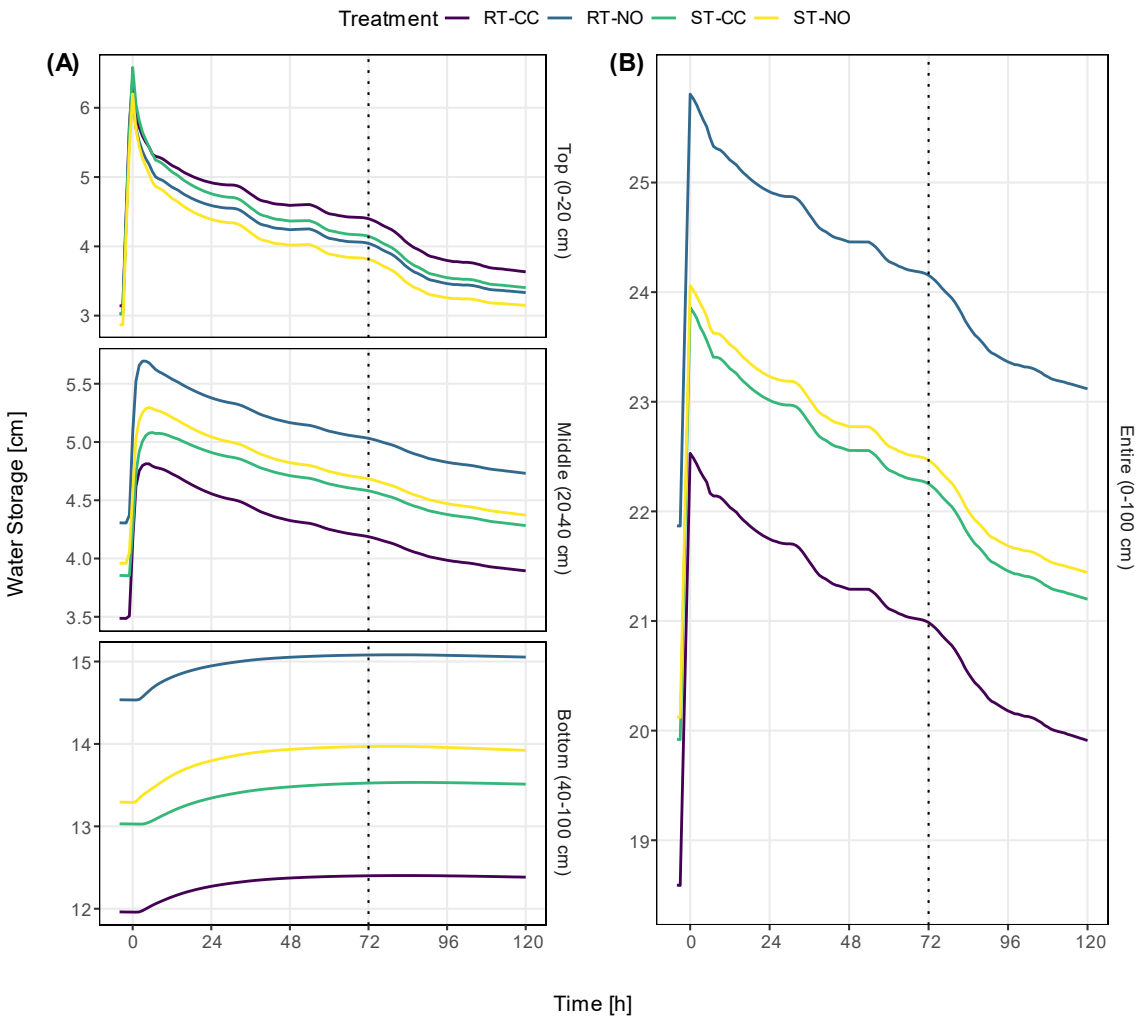

Figure A0-7 Water storage changes following 4 cm irrigation (A) for the three layers and (B) for the entire profile. Dotted vertical line indicated 3rd day after irrigation.

Tukey's HSD test comparison of means for soil hydraulic properties

Variable units for the Tukey's HSD test comparison table.

Variable name:	Units:
ρ_b	g cm^{-3}
θ (-33 kPa)	cm^3 cm^{-3}
θ (-10 kPa)	cm^3 cm^{-3}
K (-10 kPa)	$\log_{10}(cm\ d^{-1})$
K_s	$\log_{10}(cm\ d^{-1})$
PAW	cm^3 cm^{-3}
φ	cm^3 cm^{-3}
θ (-1500 kPa)	cm^3 cm^{-3}
θ (3 days)	Cm cm^{-1}
Storage (3 days)	cm

*Tukey's HSD comparison of means. Significance symbols are as follows: *** = $p < 0.05$, ** = $p < 0.1$, * = $p < 0.15$, . = $p < 0.2$. LCL and UCL are lower and upper control intervals, respectively.*

Variable	Depth	Comparison	Difference	p-value	Sig.	LCL	UCL
ρ_b	Top (0-5cm)	RT-CC -- RT-NO	-0.1	0.4274		-0.2467	0.046705
ρ_b	Top (0-5cm)	RT-CC -- ST-CC	-0.0125	0.9971		-0.1592	0.134205
ρ_b	Top (0-5cm)	RT-CC -- ST-NO	0.0225	0.984		-0.1242	0.169205
ρ_b	Top (0-5cm)	RT-NO -- ST-CC	0.0875	0.5351		-0.0592	0.234205
ρ_b	Top (0-5cm)	RT-NO -- ST-NO	0.1225	0.2671		-0.0242	0.269205
ρ_b	Top (0-5cm)	ST-CC -- ST-NO	0.035	0.9444		-0.1117	0.181705
ρ_b	Subsurface (20-25cm)	RT-CC -- RT-NO	-0.055	0.5447		-0.14835	0.038352
ρ_b	Subsurface (20-25cm)	RT-CC -- ST-CC	0	1		-0.09335	0.093352
ρ_b	Subsurface (20-25cm)	RT-CC -- ST-NO	-0.0325	0.8511		-0.12585	0.060852
ρ_b	Subsurface (20-25cm)	RT-NO -- ST-CC	0.055	0.5447		-0.03835	0.148352
ρ_b	Subsurface (20-25cm)	RT-NO -- ST-NO	0.0225	0.9429		-0.07085	0.115852
ρb	Subsurface (20-25cm)	ST-CC -- ST-NO	-0.0325	0.8511		-0.12585	0.060852
θ (-33 kPa)	Top (0-5cm)	RT-CC -- RT-NO	-0.0102	0.9424		-0.05223	0.031902
θ (-33 kPa)	Top (0-5cm)	RT-CC -- ST-CC	-0.0614	0.0249	***	-0.10343	-0.01929
θ (-33 kPa)	Top (0-5cm)	RT-CC -- ST-NO	-0.035	0.2691		-0.0771	0.007036
θ (-33 kPa)	Top (0-5cm)	RT-NO -- ST-CC	-0.0512	0.0655	**	-0.09326	-0.00912
θ (-33 kPa)	Top (0-5cm)	RT-NO -- ST-NO	-0.0249	0.5421		-0.06693	0.017202
θ (-33 kPa)	Top (0-5cm)	ST-CC -- ST-NO	0.0263	0.4968		-0.01574	0.068395
θ (-33 kPa)	Subsurface (20-25cm)	RT-CC -- RT-NO	-0.0382	0.0942	**	-0.07218	-0.0042
θ (-33 kPa)	Subsurface (20-25cm)	RT-CC -- ST-CC	-0.0568	0.0105	***	-0.09079	-0.0228
θ (-33 kPa)	Subsurface (20-25cm)	RT-CC -- ST-NO	-0.0344	0.1442	*	-0.06835	-3.61E-04
θ (-33 kPa)	Subsurface (20-25cm)	RT-NO -- ST-CC	-0.0186	0.6007		-0.0526	0.015391
θ (-33 kPa)	Subsurface (20-25cm)	RT-NO -- ST-NO	0.0038	0.9935		-0.03016	0.037829
θ (-33 kPa)	Subsurface (20-25cm)	ST-CC -- ST-NO	0.0224	0.4536		-0.01155	0.056433
θ (-10 kPa)	Top (0-5cm)	RT-CC -- RT-NO	-0.0102	0.9712		-0.0642	0.043852
θ (-10 kPa)	Top (0-5cm)	RT-CC -- ST-CC	-0.0469	0.2398		-0.10088	0.00717
θ (-10 kPa)	Top (0-5cm)	RT-CC -- ST-NO	-0.0427	0.3089		-0.09671	0.01134
θ (-10 kPa)	Top (0-5cm)	RT-NO -- ST-CC	-0.0367	0.4306		-0.09071	0.017343
θ (-10 kPa)	Top (0-5cm)	RT-NO -- ST-NO	-0.0325	0.528		-0.08654	0.021513
θ (-10 kPa)	Top (0-5cm)	ST-CC -- ST-NO	0.0042	0.9979		-0.04986	0.058195
θ (-10 kPa)	Subsurface (20-25cm)	RT-CC -- RT-NO	-0.0236	0.5638		-0.06465	0.017455

Variable	Depth	Comparison	Difference	p-value	Sig.	LCL	UCL
θ (-10 kPa)	Subsurface (20-25cm)	RT-CC -- ST-CC	-0.0477	0.0815	**	-0.08871	-0.00661
θ (-10 kPa)	Subsurface (20-25cm)	RT-CC -- ST-NO	-0.0304	0.361		-0.07142	0.010687
θ (-10 kPa)	Subsurface (20-25cm)	RT-NO -- ST-CC	-0.0241	0.5487		-0.06511	0.016991
θ (-10 kPa)	Subsurface (20-25cm)	RT-NO -- ST-NO	-0.0068	0.9803		-0.04782	0.034284
θ (-10 kPa)	Subsurface (20-25cm)	ST-CC -- ST-NO	0.0173	0.7668		-0.02376	0.058345
K (-10 kPa)	Top (0-5cm)	RT-CC -- RT-NO	0.0006	0.7252		-6.73E-04	0.001792
K (-10 kPa)	Top (0-5cm)	RT-CC -- ST-CC	-0.0011	0.228		-0.00232	1.45E-04
K (-10 kPa)	Top (0-5cm)	RT-CC -- ST-NO	-0.0007	0.5351		-0.00197	4.98E-04
K (-10 kPa)	Top (0-5cm)	RT-NO -- ST-CC	-0.0016	0.0407	***	-0.00288	-4.14E-04
K (-10 kPa)	Top (0-5cm)	RT-NO -- ST-NO	-0.0013	0.1244	*	-0.00253	-6.21E-05
K (-10 kPa)	Top (0-5cm)	ST-CC -- ST-NO	0.0004	0.91		-8.81E-04	0.001585
K (-10 kPa)	Subsurface (20-25cm)	RT-CC -- RT-NO	0.0002	0.9987		-0.00342	0.003908
K (-10 kPa)	Subsurface (20-25cm)	RT-CC -- ST-CC	0.0002	0.9991		-0.00346	0.003877
K (-10 kPa)	Subsurface (20-25cm)	RT-CC -- ST-NO	0.0004	0.992		-0.00322	0.004109
K (-10 kPa)	Subsurface (20-25cm)	RT-NO -- ST-CC	0	1		-0.0037	0.003635
K (-10 kPa)	Subsurface (20-25cm)	RT-NO -- ST-NO	0.0002	0.9992		-0.00347	0.003868
K (-10 kPa)	Subsurface (20-25cm)	ST-CC -- ST-NO	0.0002	0.9988		-0.00343	0.003899
K_s	Top (0-5cm)	RT-CC -- RT-NO	0.7535	0.0497	***	0.167864	1.339156
K_s	Top (0-5cm)	RT-CC -- ST-CC	0.338	0.5607		-0.24767	0.92362
K_s	Top (0-5cm)	RT-CC -- ST-NO	0.5701	0.1652	.	-0.01552	1.155768
K_s	Top (0-5cm)	RT-NO -- ST-CC	-0.4155	0.3946		-1.00118	0.17011
K_s	Top (0-5cm)	RT-NO -- ST-NO	-0.1834	0.8858		-0.76903	0.402258
K_s	Top (0-5cm)	ST-CC -- ST-NO	0.2321	0.797		-0.3535	0.817794
K_s	Subsurface (20-25cm)	RT-CC -- RT-NO	0.6334	0.3668		-0.2292	1.496005
K_s	Subsurface (20-25cm)	RT-CC -- ST-CC	1.0094	0.0786	**	0.146834	1.872036
K_s	Subsurface (20-25cm)	RT-CC -- ST-NO	0.9008	0.1273	*	0.038175	1.763377
K_s	Subsurface (20-25cm)	RT-NO -- ST-CC	0.376	0.7483		-0.48657	1.238632
K_s	Subsurface (20-25cm)	RT-NO -- ST-NO	0.2674	0.8888		-0.59523	1.129973
K_s	Subsurface (20-25cm)	ST-CC -- ST-NO	-0.1087	0.991		-0.97126	0.753942
PAW	Top (0-5cm)	RT-CC -- RT-NO	-0.0279	0.5792		-0.07745	0.021636
PAW	Top (0-5cm)	RT-CC -- ST-CC	-0.0619	0.0576	**	-0.11143	-0.01235
PAW	Top (0-5cm)	RT-CC -- ST-NO	-0.0656	0.0427	***	-0.11517	-0.01608
PAW	Top (0-5cm)	RT-NO -- ST-CC	-0.034	0.4223		-0.08352	0.015561
PAW	Top (0-5cm)	RT-NO -- ST-NO	-0.0377	0.3379		-0.08726	0.011823
PAW	Top (0-5cm)	ST-CC -- ST-NO	-0.0037	0.998		-0.05328	0.045805
PAW	Subsurface (20-25cm)	RT-CC -- RT-NO	0.0015	0.9998		-0.0423	0.045319
PAW	Subsurface (20-25cm)	RT-CC -- ST-CC	-0.0317	0.3777		-0.07555	0.012068
PAW	Subsurface (20-25cm)	RT-CC -- ST-NO	-0.0107	0.94		-0.05456	0.03306
PAW	Subsurface (20-25cm)	RT-NO -- ST-CC	-0.0333	0.3403		-0.07706	0.010557
PAW	Subsurface (20-25cm)	RT-NO -- ST-NO	-0.0123	0.9147		-0.05607	0.031549
PAW	Subsurface (20-25cm)	ST-CC -- ST-NO	0.021	0.6921		-0.02282	0.0648
φ	Top (0-5cm)	RT-CC -- RT-NO	0.0375	0.4772		-0.02098	0.095982
φ	Top (0-5cm)	RT-CC -- ST-CC	0.005	0.9971		-0.05348	0.063482
φ	Top (0-5cm)	RT-CC -- ST-NO	-0.01	0.9781		-0.06848	0.048482
φ	Top (0-5cm)	RT-NO -- ST-CC	-0.0325	0.5893		-0.09098	0.025982
φ	Top (0-5cm)	RT-NO -- ST-NO	-0.0475	0.2877		-0.10598	0.010982
φ	Top (0-5cm)	ST-CC -- ST-NO	-0.015	0.9324		-0.07348	0.043482
φ	Subsurface (20-25cm)	RT-CC -- RT-NO	0.02	0.5729		-0.01522	0.055216
φ	Subsurface (20-25cm)	RT-CC -- ST-CC	-0.0025	0.9983		-0.03772	0.032716
φ	Subsurface (20-25cm)	RT-CC -- ST-NO	0.01	0.9114		-0.02522	0.045216
φ	Subsurface (20-25cm)	RT-NO -- ST-CC	-0.0225	0.4801		-0.05772	0.012716
φ	Subsurface (20-25cm)	RT-NO -- ST-NO	-0.01	0.9114		-0.04522	0.025216
φ	Subsurface (20-25cm)	ST-CC -- ST-NO	0.0125	0.8439		-0.02272	0.047716
θ (-1500 kPa)	Top (0-5cm)	RT-CC -- RT-NO	0.0177	0.6108		-0.01513	0.050608
θ (-1500 kPa)	Top (0-5cm)	RT-CC -- ST-CC	0.0005	1		-0.03234	0.033397
θ (-1500 kPa)	Top (0-5cm)	RT-CC -- ST-NO	0.0306	0.1925	.	-0.00227	0.063462
θ (-1500 kPa)	Top (0-5cm)	RT-NO -- ST-CC	-0.0172	0.6326		-0.05008	0.015657

Variable	Depth	Comparison	Difference	p-value	Sig.	LCL	UCL
θ (-1500 kPa)	Top (0-5cm)	RT-NO -- ST-NO	0.0129	0.8033		-0.02001	0.045721
θ (-1500 kPa)	Top (0-5cm)	ST-CC -- ST-NO	0.0301	0.2037		-0.0028	0.062932
θ (-1500 kPa)	Subsurface (20-25cm)	RT-CC -- RT-NO	-0.0397	0.3701		-0.09399	0.014583
θ (-1500 kPa)	Subsurface (20-25cm)	RT-CC -- ST-CC	-0.0251	0.7154		-0.07934	0.029231
θ (-1500 kPa)	Subsurface (20-25cm)	RT-CC -- ST-NO	-0.0236	0.7497		-0.07789	0.030678
θ (-1500 kPa)	Subsurface (20-25cm)	RT-NO -- ST-CC	0.0146	0.9226		-0.03964	0.068933
θ (-1500 kPa)	Subsurface (20-25cm)	RT-NO -- ST-NO	0.0161	0.9008		-0.03819	0.07038
θ (-1500 kPa)	Subsurface (20-25cm)	ST-CC -- ST-NO	0.0014	0.9999		-0.05284	0.055732
θ (3 days)	Top (0-20cm)	RT-CC -- RT-NO	0.0074	0.7064		-0.00836	0.023099
θ (3 days)	Top (0-20cm)	RT-CC -- ST-CC	0.023	0.0244	***	0.00728	0.038742
θ (3 days)	Top (0-20cm)	RT-CC -- ST-NO	0.0202	0.0505	**	0.004442	0.035905
θ (3 days)	Top (0-20cm)	RT-NO -- ST-CC	0.0156	0.1531	.	-8.83E-05	0.031374
θ (3 days)	Top (0-20cm)	RT-NO -- ST-NO	0.0128	0.2861		-0.00293	0.028537
θ (3 days)	Top (0-20cm)	ST-CC -- ST-NO	-0.0028	0.9745		-0.01857	0.012894
Storage (3 days)	Top (0-20cm)	RT-CC -- RT-NO	0.1326	0.7064		-0.15054	0.415789
Storage (3 days)	Top (0-20cm)	RT-CC -- ST-CC	0.4142	0.0244	***	0.131036	0.697364
Storage (3 days)	Top (0-20cm)	RT-CC -- ST-NO	0.3631	0.0505	**	0.079961	0.646289
Storage (3 days)	Top (0-20cm)	RT-NO -- ST-CC	0.2816	0.1531	.	-0.00159	0.564739
Storage (3 days)	Top (0-20cm)	RT-NO -- ST-NO	0.2305	0.2861		-0.05266	0.513664
Storage (3 days)	Top (0-20cm)	ST-CC -- ST-NO	-0.0511	0.9745		-0.33424	0.232089
θ (3 days)	Middle (20-40cm)	RT-CC -- RT-NO	0.0041	0.9846		-0.02286	0.031
θ (3 days)	Middle (20-40cm)	RT-CC -- ST-CC	-0.0147	0.6047		-0.04158	0.012272
θ (3 days)	Middle (20-40cm)	RT-CC -- ST-NO	-0.0038	0.9873		-0.03075	0.02311
θ (3 days)	Middle (20-40cm)	RT-NO -- ST-CC	-0.0187	0.4109		-0.04566	0.008199
θ (3 days)	Middle (20-40cm)	RT-NO -- ST-NO	-0.0079	0.9038		-0.03482	0.019038
θ (3 days)	Middle (20-40cm)	ST-CC -- ST-NO	0.0108	0.7897		-0.01609	0.037767
Storage (3 days)	Middle (20-40cm)	RT-CC -- RT-NO	0.0814	0.9846		-0.45711	0.620008
Storage (3 days)	Middle (20-40cm)	RT-CC -- ST-CC	-0.2931	0.6047		-0.83168	0.245433
Storage (3 days)	Middle (20-40cm)	RT-CC -- ST-NO	-0.0764	0.9873		-0.61491	0.462208
Storage (3 days)	Middle (20-40cm)	RT-NO -- ST-CC	-0.3746	0.4109		-0.91313	0.163983
Storage (3 days)	Middle (20-40cm)	RT-NO -- ST-NO	-0.1578	0.9038		-0.69636	0.380758
Storage (3 days)	Middle (20-40cm)	ST-CC -- ST-NO	0.2168	0.7897		-0.32178	0.755333
θ (3 days)	Bottom (40-100cm)	RT-CC -- RT-NO	-0.0017	0.9276		-0.00813	0.004745
θ (3 days)	Bottom (40-100cm)	RT-CC -- ST-CC	-0.0009	0.9875		-0.00735	0.005531
θ (3 days)	Bottom (40-100cm)	RT-CC -- ST-NO	-0.0038	0.5361		-0.01027	0.002603
θ (3 days)	Bottom (40-100cm)	RT-NO -- ST-CC	0.0008	0.9917		-0.00565	0.007224
θ (3 days)	Bottom (40-100cm)	RT-NO -- ST-NO	-0.0021	0.8671		-0.00858	0.004297
θ (3 days)	Bottom (40-100cm)	ST-CC -- ST-NO	-0.0029	0.7242		-0.00937	0.003511
Storage (3 days)	Bottom (40-100cm)	RT-CC -- RT-NO	-0.105	0.9276		-0.50416	0.294159
Storage (3 days)	Bottom (40-100cm)	RT-CC -- ST-CC	-0.0562	0.9875		-0.45541	0.342909
Storage (3 days)	Bottom (40-100cm)	RT-CC -- ST-NO	-0.2378	0.5361		-0.63691	0.161409
Storage (3 days)	Bottom (40-100cm)	RT-NO -- ST-CC	0.0488	0.9917		-0.35041	0.447909
Storage (3 days)	Bottom (40-100cm)	RT-NO -- ST-NO	-0.1328	0.8671		-0.53191	0.266409
Storage (3 days)	Bottom (40-100cm)	ST-CC -- ST-NO	-0.1815	0.7242		-0.58066	0.217659
θ (3 days)	Entire (0-100cm)	RT-CC -- RT-NO	0.0011	0.1847	.	-6.64E-05	0.002241
θ (3 days)	Entire (0-100cm)	RT-CC -- ST-CC	0.0007	0.5762		-5.01E-04	0.001806
θ (3 days)	Entire (0-100cm)	RT-CC -- ST-NO	0.0005	0.7652		-6.66E-04	0.001641
θ (3 days)	Entire (0-100cm)	RT-NO -- ST-CC	-0.0004	0.8195		-0.00159	7.19E-04
θ (3 days)	Entire (0-100cm)	RT-NO -- ST-NO	-0.0006	0.6375		-0.00175	5.54E-04
θ (3 days)	Entire (0-100cm)	ST-CC -- ST-NO	-0.0002	0.9869		-0.00132	9.89E-04
Storage (3 days)	Entire (0-100cm)	RT-CC -- RT-NO	0.1087	0.1847	.	-0.00664	0.22414
Storage (3 days)	Entire (0-100cm)	RT-CC -- ST-CC	0.0652	0.5762		-0.05014	0.18064
Storage (3 days)	Entire (0-100cm)	RT-CC -- ST-NO	0.0487	0.7652		-0.06664	0.16414
Storage (3 days)	Entire (0-100cm)	RT-NO -- ST-CC	-0.0435	0.8195		-0.15889	0.07189
Storage (3 days)	Entire (0-100cm)	RT-NO -- ST-NO	-0.06	0.6375		-0.17539	0.05539
Storage (3 days)	Entire (0-100cm)	ST-CC -- ST-NO	-0.0165	0.9869		-0.13189	0.09889

Appendix C: Supplemental Figures to Chapter 4

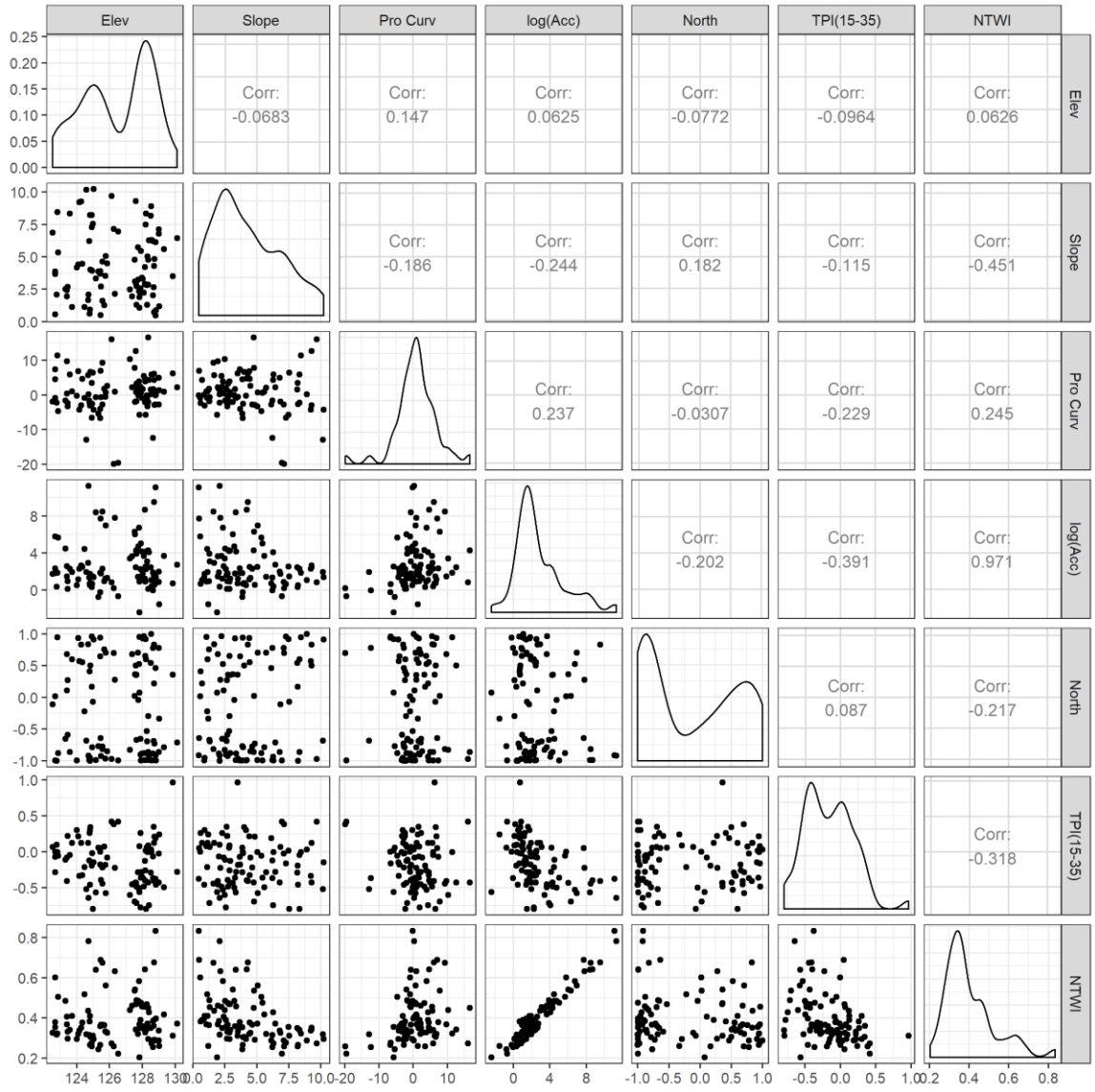

Figure A0-8 Plot matrix of selected terrain variables of the ground sampling points (from 1 m resolution raster). Top right diagonal panels show Pearson's correlation value, bottom left diagonal panels show scatter plot, and diagonal panels show density plots of variables. NTWI is the normalized topographic wetness index calculated from flow accumulation raster generated by multi-flow algorithm and TPI (15,35) is topographic position index with an inner diameter of 15 m and an outer diameter of 35 m.

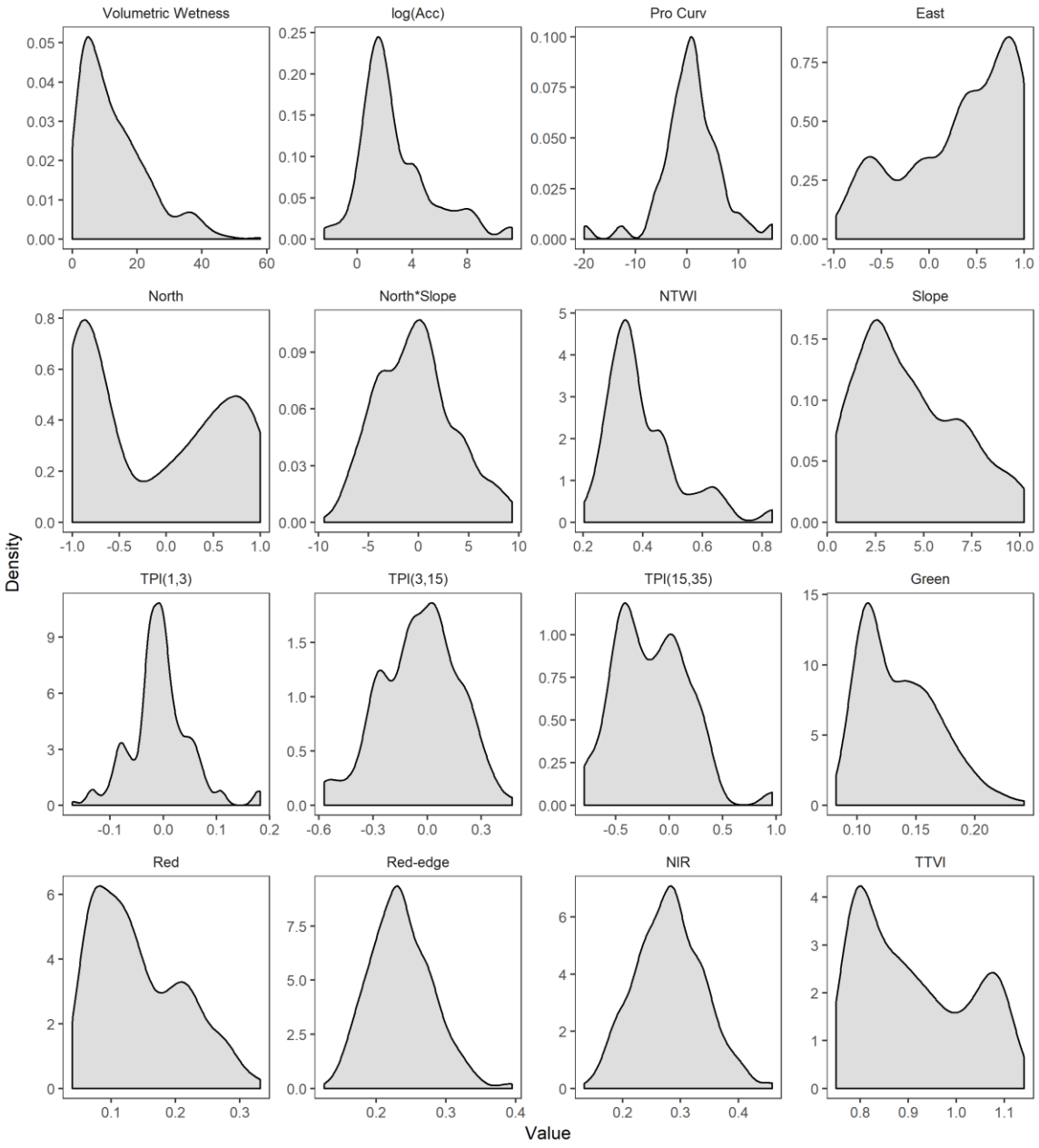

Figure A0-9 Kernel density estimate of values from selected variables in the data.

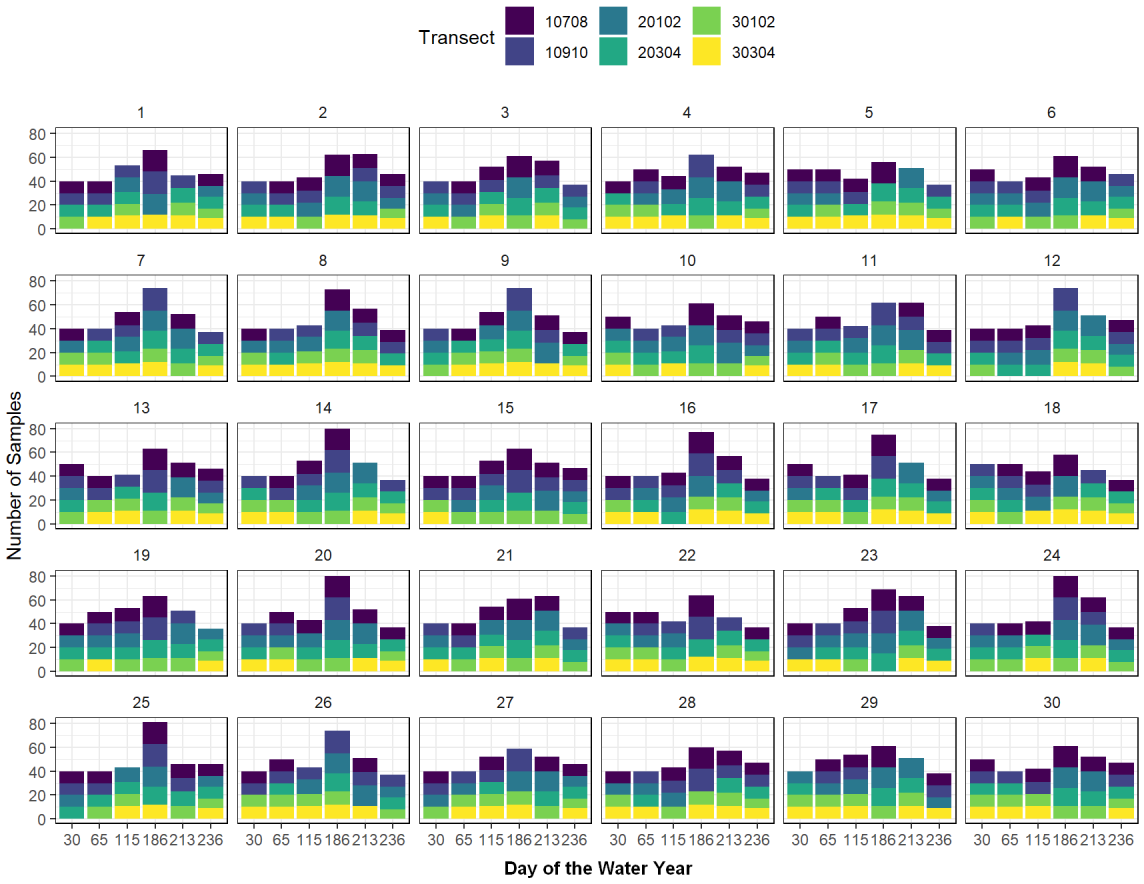

Figure A0-10 Number of samples in the 30 unique training sets grouped by sampling day and transect.

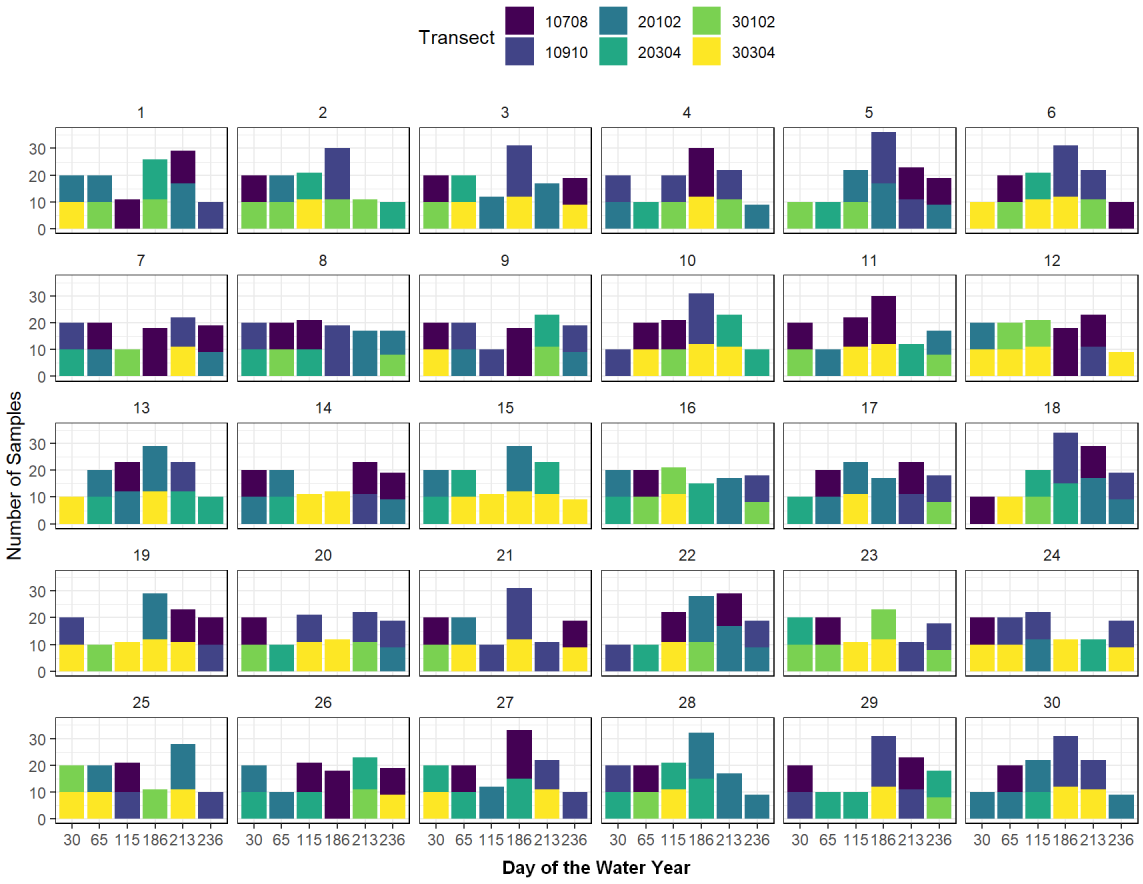

Figure A0-11 Number of samples in the 30 unique testing sets grouped by sampling day and transect.

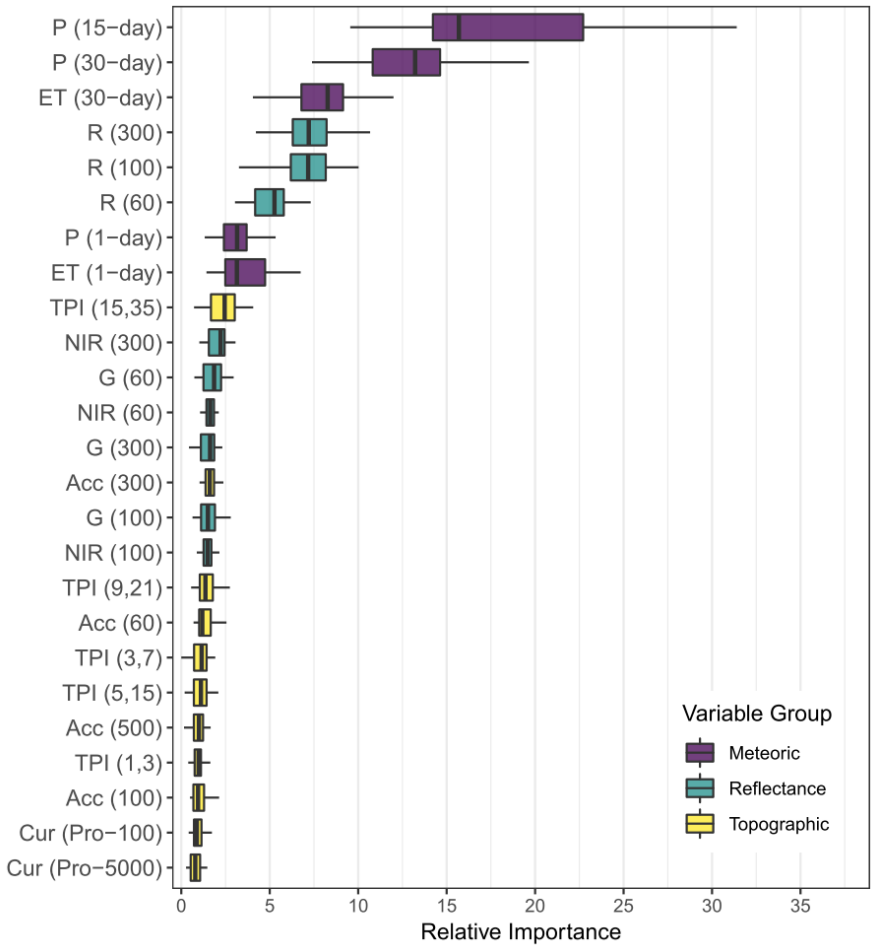

Figure A0-12 Sum of the relative variable importance distribution of the top 25 most important variables.